

REFERENCE
AUTHORS

1 (bases 1 to 196514)
 Allen, D., Marie, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Muzny, D., H. H., Albrooks, S., Amin, A., Anquiano, D.,
 Ayalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Herrandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowitz, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, W., L., Louiseged, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Manum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, R.,
 Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okwuonu, G.,
 Olatunmbun, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L., Puazo, M., Quairo, J., Rachlin, E.,
 Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, R., Richards, S., Riggs, F., Rivers, C., Rodkey, I., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C. D., Sma, J., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tinney, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D. R.,
 Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
 2 (bases 1 to 196514)
 Worley, K. C.
 Direct Submission
 Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 196514)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 8, 2002 this sequence version replaced gi:21628958.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

 Center project name: GCKN
 Center clone name: CH230-7L23
 ----- Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 136389 bases at least Q40
 Consensus quality: 141908 bases at least Q30
 Consensus quality: 145885 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 70 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence,
 as soon as it is available and the accession number will
 be preserved.

1 1183: contig of 1183 bp in length
 * 1184 1283: gap of unknown length
 * 1284 2907: contig of 1624 bp in length
 * 2908 3007: gap of unknown length
 * 3008 4411: contig of 1434 bp in length
 * 4412 4541: gap of unknown length
 * 4542 5727: contig of 1186 bp in length
 * 5728 5827: gap of unknown length
 * 5828 6853: contig of 1026 bp in length
 * 6854 6953: gap of unknown length
 * 6954 7970: contig of 1017 bp in length
 * 7971 8070: gap of unknown length
 * 8071 9653: contig of 1583 bp in length
 * 9654 11015: contig of 1262 bp in length
 * 9754 11115: gap of unknown length
 * 11116 12358: contig of 1243 bp in length
 * 12359 12458: gap of unknown length
 * 12459 13782: contig of 1324 bp in length
 * 13783 13882: gap of unknown length
 * 13883 15001: contig of 1119 bp in length
 * 15002 15101: gap of unknown length
 * 15102 16162: contig of 1061 bp in length
 * 16163 16262: gap of unknown length
 * 16263 17807: contig of 1545 bp in length
 * 17808 17907: gap of unknown length
 * 17908 19137: contig of 1230 bp in length
 * 19138 19237: gap of unknown length
 * 19238 20680: contig of 1443 bp in length
 * 20681 22095: contig of 1315 bp in length
 * 22096 22195: gap of unknown length
 * 22196 24074: contig of 1879 bp in length
 * 24075 24174: gap of unknown length
 * 24175 25334: contig of 1160 bp in length
 * 25335 26813: contig of 1379 bp in length
 * 26814 26913: gap of unknown length
 * 26914 28582: contig of 1669 bp in length
 * 28583 28682: gap of unknown length
 * 28683 30029: contig of 1347 bp in length
 * 30030 31734: contig of 1605 bp in length
 * 31735 31834: gap of unknown length
 * 31835 34518: contig of 2684 bp in length
 * 34519 34618: gap of unknown length
 * 34619 36561: contig of 1943 bp in length
 * 36562 38272: contig of 1611 bp in length
 * 38273 38372: gap of unknown length
 * 38373 39819: contig of 1447 bp in length
 * 39820 39919: gap of unknown length
 * 39920 41753: contig of 1834 bp in length
 * 41754 41854: gap of unknown length
 * 41854 44464: contig of 2611 bp in length
 * 44465 44564: gap of unknown length

	ORGANISM	Rattus norvegicus
	MRNA.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
	REFERENCE	1 (bases 1 to 1161)
	AUTHORS	Suga, H. and Takao, K.
	TITLE	Expression of the rat SENR in the urinary bladder tissues
	JOURNAL	Published Only in Database (1998)
	REFERENCE	2 (bases 1 to 1161)
	AUTHORS	Suga, H. and Takao, K.
	TITLE	Direct Submission
	JOURNAL	Submitted (16-MAR-1998) Hinako Suga, Inst. for Mol. & Cell. Regulation, Gunma Univ., Laboratory of Peptide and Protein Research; 3-39-15 Showa-machi, Maebashi, Gunma 371-8512, Japan (E-mail:hsgu@news.sb.gunma-u.ac.jp, Tel:81-27-220-8876, Fax:81-27-220-8849)
	FEATURES	Location/Qualifiers 1..1161
	source	/organism="Rattus norvegicus" /strain="Wistar" /db_xref="taxon:10116" /sex="male" /tissue_type="urinary bladder"
	CDS	1..1161 /codon_start=1 /product="SENR" /protein_id="BAA25251.1" /db_xref="GI:2981624."
	BASE COUNT	205 a 381 c 301 g 274 t
	ORIGIN	LQSSCHSPSPGSGFLPSRVHLQQDSSGRSLSSSQCATETMLSPVPRNGALL*
	Query Match	Score 43.6; DB 10; Length 1161;
	Best Local Similarity	51.2%; Pred. No. 0.25;
	Matches 126; Conservative	0; Mismatches 119; Indels 1; Gaps 1;
QY	105	GGACCCAGTCAGATGCCAGTCTCCTCACCGGGCTTGAGATGTCACCAAGTCTCCTCTCTCCA 164
Db	574	GGACCAGCTGGATGGCAAGCATCATGGTAGGGTCAGCAGCAGTGCACCAAGTGTGC 515
QY	165	CCGAGGGCCACACCCCCTGGAGACCCCTTGAGTCAGTCTTTGCTTGCATCTCTTT 224
Db	514	CCAGCACACAGCATCTACGGTAACCCCTTGAGCGCTGGACTGTGTC-AGAGCCCTCAGT 456
QY	225	CTGGCCCTCATAGTGGGGCTTGCCCATTTGCCCTTCACTCCAGATCTCTCTTTCAGTGCC 284
Db	455	ACGGCTGCATAGGTTTCGGTGTCTATTATGGTCAGGGTGAAGATGCTGSCGTGCTTGTG 396
QY	285	AGGAAGTCATCTTGACITTAACCTTCCAGACCCCCCTTCAGTTTTCCAGTCCTTAGAG 344
Db	395	AGGAAGTCAGGTAAGAAGAGACTCTGCAGCCACATCTCCAAAGTGCAGTCCTTAGTG 336
QY	345	AGGTGG 350
Db	335	ACGTAG 330
RESULT 8		
AB029611/c		
LOCUS	AB029611	1161 bp mRNA linear ROD 10-JUL-1999
DEFINITION	Rattus norvegicus SENR/GP14 mRNA for G protein-coupled receptor,	
complete cds.		
ACCESSION	AB029611	
VERSION	AB029611.1	GI:5441249
KEYWORDS	SENR/GP14; G protein-coupled receptor.	
SOURCE	Rattus norvegicus pheochromocytoma cell_line:PC 12 cdna to mrna.	
* 44565	45610:	contig of 2046 bp in length
* 46611	46710:	gap of unknown length
* 48567	48567:	contig of 1857 bp in length
* 48568	48567:	gap of unknown length
* 50569	50568:	contig of 1901 bp in length
* 50669	50668:	gap of unknown length
* 51872	51872:	contig of 1204 bp in length
* 51972	51972:	gap of unknown length
* 53621	53621:	contig of 1649 bp in length
* 53622	53721:	gap of unknown length
* 53722	56728:	contig of 3007 bp in length
* 56828	56828:	gap of unknown length
* 56829	58508:	contig of 1680 bp in length
* 58509	58508:	gap of unknown length
* 58609	60295:	contig of 1887 bp in length
* 60396	60395:	gap of unknown length
* 60396	62519:	contig of 2124 bp in length
* 62520	62619:	gap of unknown length
* 62620	65449:	contig of 2830 bp in length
* 65450	65549:	gap of unknown length
* 65550	68064:	contig of 2515 bp in length
* 68065	68164:	gap of unknown length
* 68165	70481:	contig of 2317 bp in length
* 70482	70581:	gap of unknown length
* 70582	73499:	contig of 2918 bp in length
* 73500	73599:	gap of unknown length
* 73600	76013:	contig of 2414 bp in length
* 76014	76113:	gap of unknown length
* 76114	79286:	contig of 3173 bp in length
* 79287	79386:	gap of unknown length
* 79387	81270:	contig of 1884 bp in length
* 81271	81370:	gap of unknown length
* 81371	85687:	contig of 4317 bp in length
* 85688	85787:	gap of unknown length
* 85788	88465:	contig of 2678 bp in length
* 88466	88565:	gap of unknown length
* 88566	92552:	contig of 3987 bp in length
* 92553	92652:	gap of unknown length
* 92653	95618:	contig of 2966 bp in length
* 95619	95718:	gap of unknown length
* 95719	99578:	contig of 3860 bp in length
* 99579	99678:	gap of unknown length
* 99679	102566:	contig of 2888 bp in length
Query Match	6.2%; Score 46.4; DB 2; Length 196514;	
Best Local Similarity	52.0%; Pred. No. 0.035;	
Matches 104; Conservative	0; Mismatches 96; Indels 0; Gaps 0;	
QY	159	TCTCCACGAGGCGCACCCCTGGAGACCCCTTGAGTCAGTCTTTGCTTGCATCA 218
Db	93628	TATGACCAAGTTCGCTGTGGCCCTGCATCTCTATACTTTGGGGTCTCTTTTATGA 93569
QY	219	CTCTTTCTGGCTCATAGTGGGGCTGGCCATTTGCCCTTCACTCCAGATCTCTCTTTC 278
Db	93568	CGCCAGCTGGGTCACCCCTGAAGCTCTAGTAATGCACTTGGCAGCATGTAGACCTTCTC 93509
QY	279	AGTCCAGGAGTGCATCTTGACITTAACCTTCCAGACCCCCCTTCAGTTTCCAGTCC 338
Db	93508	AGGGGAGCATTAAGCTCCAGATTAAACTCTTCGCTTCCCTCCCTCGGTCGTGAAGCC 93449
QY	339	TTAGAGAGTGGACTTCTGA 358
Db	93448	TGGAAGAGTGGAGGCTGA 93429
RESULT 7		
AB012210/c		
LOCUS	AB012210	1161 bp mRNA linear ROD 24-MAR-1998
DEFINITION	Rattus norvegicus mRNA for SENR, complete cds.	
ACCESSION	AB012210	
VERSION	AB012210.1	GI:2981623
KEYWORDS</		

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Liu, H., Zou, M., Suga, H. and Takao, K.
TITLE The SENR/GPR14 expresses in rat pheochromocytoma PC 12 cells
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 1161)
AUTHORS Liu, H., Zou, M., Suga, H. and Takao, K.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-1999) Hui-Yu Liu, The University of Kansas,
Pharmacology and Toxicology; 5064 Mallott Hall, Lawrence, Kansas
66047-2505, USA (E-mail:huiyuliueagle.cc.ukans.edu,
URL:http://129.237.102.160/pharmtox/index.html, tel:785-864-3538,
Fax:785-864-5219)
FEATURES Location/Qualifiers
source 1..1161
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="PC 12"
/tissue_type="pheochromocytoma"
gene 1..1161
/gene="SENR/GPR14"
CDS 1..1161
/gene="SENR/GPR14"
/codon_start=1
/product="G protein-coupled receptor"
/protein_id="BA82357.1"
/translation="GI:5441250"
LKDVLATVIGAVLGMGVGVNVTLVNCRFLRASNYVYVNLALDLYLL
STPIIATVYKDMDFGVDGCVLFSLDFLMHSAIFITIMSSERYAAVRLDITVQ
RSGYRKLVLVLTWLLALLLPMLAIQLVRGSKSLCLPAMGPHRAHRTYLLFGT
SVGGLVIGLVRLARAYWLSSQASPKQRLPNRVLVLLIGLVLLFWACILPEW
LWLLAOYHEAPLPETARIYNYLTCLTGNSCINFEFLTLTKNRYELRGQR
LSSCHSPSPSPGSPFSPRVHLOQDSGRSLSSSQOATETLMLSPVPRNGALL"
BASE COUNT 205 a 381 c 301 g 274 t
ORIGIN
Query Match 5.8%; Score 43.6; DB 10; Length 1161;
Best Local Similarity 51.2%; Pred. No. 0.25;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
Qy 105 GGACAGTGCAGATGCCACGCTCTCCACGGGCTCAGAATGCTCACAGCTTCTCTCTCCA 164
Db 574 GGACAGCTGGATGGCAAGCATCATGGTAGGTGCTCAGCAGCAGTGCACCAACCGAGTGC 515
Qy 165 CCGAGGGCCACAGCCCTTCAGACCCCTTCAGCTGAGTGCCTTGTCTTGCATACCTCTT 224
Db 514 CCAGCACCAGCAGCTTACGGTAACCCCTTGAGCGCTGGAGTGTGCTCC-AGAGCCCTCAGT 456
Qy 225 CTGGCCTCATAGTGGGCTTGGCCATGTCCTTCACTCCAGATCTCTCTTTCAGGTCC 284
Db 455 ACGGCTCATAGCTTCGCTGCTCATTATGTCAGGGTGAAGATGCTGGCTGCATTGTC 396
Qy 285 AGGAAGTGCACTTTGAACCTTAACCTTCCAGACCCCTTCAGTCTTTTCCAGTCCCTTAGAG 344
Db 395 AGGAAGTCCAGGCTTAAAGAGGACTCTGCAGCCACATCTCCAAAGTCCAGTCTCTTAGTG 336
Qy 345 AGGTGG 350
Db 335 ACGTAG 330
RESULT 9
AR116596/c
LOCUS AR116596 1539 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 6 from patent US 6133420.
ACCESSION AR116596
VERSION AR116596.1 GI:14096918
KEYWORDS
SOURCE Unknown.
REFERENCE 1 (bases 1 to 1539)
AUTHORS Aiyar, N.V., Ames, R.S., Arnold, A. Romanic., Al-Barazanji, K.,
Bergsma, D.J., Chambers, J., Douglas, S.A., Foley, J.J., Gout, B.,
Khandoudi, N., Sarau, H.M., Shabon, U. and Willette, R.N.
TITLE Method of finding agonist and antagonist to human and rat GPR14
JOURNAL
FEATURES Location/Qualifiers
source 1..1539
/organism="unknown"
BASE COUNT 291 a 480 c 405 g 363 t
ORIGIN
Query Match 5.8%; Score 43.6; DB 6; Length 1539;
Best Local Similarity 51.2%; Pred. No. 0.25;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
Qy 105 GGACAGTGCAGATGCCACGCTCTCTCAGGGGCTCAGAATGCTCACAGCTTCTCTCTCCA 164
Db 854 GGACAGCTGGATGGCAAGCATCATGGTAGGTGCTCAGCAGCAGTGCACCAACCGAGTGC 795
Qy 165 CCGAGGGCCACAGCCCTTCAGACCCCTTCAGCTGAGTGCCTTGTCTTGCATACCTCTT 224
Db 794 CCAGCACCAGCAGCTTACGGTAACCCCTTGAGCGCTGGAGTGTGCTCC-AGAGCCCTCAGT 736
Qy 225 CTGGCCTCATAGTGGGCTTGGCCATGTCCTTCACTCCAGATCTCTCTCTTTCAGGTCC 284

ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1539)
AUTHORS Ames, R.S., Jr., Sarau, H.M., Foley, J.J., Shabon, U., Bergsma, D. and
Chambers, J.K.
TITLE GPR14 polypeptides
JOURNAL Patent: US 6133420-A 6 17-OCT-2000;
FEATURES Location/Qualifiers
source 1..1539
/organism="unknown"
BASE COUNT 291 a 480 c 405 g 363 t
ORIGIN
Query Match 5.8%; Score 43.6; DB 6; Length 1539;
Best Local Similarity 51.2%; Pred. No. 0.25;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
Qy 105 GGACAGTGCAGATGCCACGCTCTCTCAGGGGCTCAGAATGCTCACAGCTTCTCTCTCCA 164
Db 854 GGACAGCTGGATGGCAAGCATCATGGTAGGTGCTCAGCAGCAGTGCACCAACCGAGTGC 795
Qy 165 CCGAGGGCCACAGCCCTTCAGACCCCTTCAGCTGAGTGCCTTGTCTTGCATACCTCTT 224
Db 794 CCAGCACCAGCAGCTTACGGTAACCCCTTGAGCGCTGGAGTGTGCTCC-AGAGCCCTCAGT 736
Qy 225 CTGGCCTCATAGTGGGCTTGGCCATGTCCTTCACTCCAGATCTCTCTCTTTCAGGTCC 284
Db 735 ACGGCTCATAGCTTCGCTGCTCATTATGTCAGGGTGAAGATGCTGGCGTGCATTGTC 676
Qy 285 AGGAAGTGCACTTTGAACCTTAACCTTCCAGACCCCTTCAGTCTTTTCCAGTCCCTTAGAG 344
Db 675 AGGAAGTCCAGGCTTAAAGAGGACTCTGCAGCCACATCTCCAAAGTCCAGTCTCTTAGTG 616
Qy 345 AGGTGG 350
Db 615 ACGTAG 610
RESULT 10
AR121155/c
LOCUS AR121155 1539 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 6 from patent US 6159700.
ACCESSION AR121155
VERSION AR121155.1 GI:14104731
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1539)
AUTHORS Aiyar, N.V., Ames, R.S., Arnold, A. Romanic., Al-Barazanji, K.,
Bergsma, D.J., Chambers, J., Douglas, S.A., Foley, J.J., Gout, B.,
Khandoudi, N., Sarau, H.M., Shabon, U. and Willette, R.N.
TITLE Method of finding agonist and antagonist to human and rat GPR14
JOURNAL
FEATURES Location/Qualifiers
source 1..1539
/organism="unknown"
BASE COUNT 291 a 480 c 405 g 363 t
ORIGIN
Query Match 5.8%; Score 43.6; DB 6; Length 1539;
Best Local Similarity 51.2%; Pred. No. 0.25;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
Qy 105 GGACAGTGCAGATGCCACGCTCTCTCAGGGGCTCAGAATGCTCACAGCTTCTCTCTCCA 164
Db 854 GGACAGCTGGATGGCAAGCATCATGGTAGGTGCTCAGCAGCAGTGCACCAACCGAGTGC 795
Qy 165 CCGAGGGCCACAGCCCTTCAGACCCCTTCAGCTGAGTGCCTTGTCTTGCATACCTCTT 224
Db 794 CCAGCACCAGCAGCTTACGGTAACCCCTTGAGCGCTGGAGTGTGCTCC-AGAGCCCTCAGT 736
Qy 225 CTGGCCTCATAGTGGGCTTGGCCATGTCCTTCACTCCAGATCTCTCTCTTTCAGGTCC 284

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 212337)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-31H15
Unpublished
2 (bases 1 to 212337)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,X., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mingna,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 212337)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mingna,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 2, 2002 this sequence version replaced gi:20336104.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20914
Center clone name: 31_H_15
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 204794 bases at least Q40

Consensus quality: 207604 bases at least Q30
Consensus quality: 208522 bases at least Q20
Insert size: 180000; agarose-fp
Quality size: 209337; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 6.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 43025: contig of 43025 bp in length
* 43026 43125: gap of 100 bp
* 43126 43927: contig of 802 bp in length
* 43928 44027: gap of 100 bp
* 44028 45186: contig of 1159 bp in length
* 45187 45286: gap of 100 bp
* 45287 46369: contig of 1083 bp in length
* 46370 46463: gap of 100 bp
* 46470 48175: contig of 1706 bp in length
* 48176 48275: gap of 100 bp
* 48276 49404: contig of 1129 bp in length
* 49405 49504: gap of 100 bp
* 49505 50867: contig of 1363 bp in length
* 50868 50967: gap of 100 bp
* 50968 53166: contig of 2199 bp in length
* 53167 53266: gap of 100 bp
* 53267 55043: contig of 1777 bp in length
* 55044 55143: gap of 100 bp
* 55144 55269: contig of 1126 bp in length
* 55270 56369: gap of 100 bp
* 56370 57913: contig of 1544 bp in length
* 57914 58013: gap of 100 bp
* 58014 58899: contig of 886 bp in length
* 58900 58999: gap of 100 bp
* 59000 60426: contig of 1427 bp in length
* 60427 60526: gap of 100 bp
* 60527 61853: contig of 1327 bp in length
* 61854 61953: gap of 100 bp
* 61954 63243: contig of 1290 bp in length
* 63244 63343: gap of 100 bp
* 63344 64724: contig of 1381 bp in length
* 64725 64824: gap of 100 bp
* 64825 66189: contig of 1365 bp in length
* 66190 66289: gap of 100 bp
* 66290 68205: contig of 1916 bp in length
* 68206 68305: gap of 100 bp
* 68306 70085: contig of 1780 bp in length
* 70086 70185: gap of 100 bp
* 70186 72184: contig of 1999 bp in length
* 72185 72284: gap of 100 bp
* 72285 75333: contig of 3049 bp in length
* 75334 75433: gap of 100 bp
* 75434 78079: contig of 2646 bp in length
* 78080 78179: gap of 100 bp
* 78180 80140: contig of 1961 bp in length
* 80141 80240: gap of 100 bp
* 80241 82033: contig of 1793 bp in length
* 82034 82133: gap of 100 bp
* 82134 85563: contig of 3430 bp in length
* 85564 85663: gap of 100 bp
* 85664 89615: contig of 3952 bp in length
* 89616 89715: gap of 100 bp
* 89716 94766: contig of 5051 bp in length
* 94767 94866: gap of 100 bp
* 94867 106620: contig of 11754 bp in length
* 106621 106720: gap of 100 bp
* 106721 120992: contig of 14272 bp in length

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:50:38 ; Search time 124.796 Seconds
(without alignments)
13552.085 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751

Sequence: 1 aaagcagactctagcactca.....gggaggtctacacgttctgc 751

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	100.0	10562	21	AAA93876 Human beta3Gal-T5
2	145.8	19.4	1745	23	AA572914 DNA encoding novel
3	43.6	5.8	1189	21	AA60761 Rat SENR ligand nu
4	43.6	5.8	1539	20	AA211175 Rat GPR14 coding s
5	43.6	5.8	1539	24	AA32925 Rattus norvegicus
6	43.6	5.8	3122	22	AA91237 Rat Urotensin II r
7	39.4	5.2	1158	22	AAC84494 Mouse urotensin-II
8	36.4	4.8	10732	21	AA10594 Gene encoding a su
9	35.6	4.7	67212	21	AAA08954 WFS1 variant genom

10	35.6	4.7	123219	23	AAH88703 Human DNA sequence
c 11	35	4.7	3695	22	AAS36579 Human cardiovascular
c 12	35	4.7	3695	22	AAS36582 Human cardiovascular
13	34.8	4.6	62944	24	ABL66947 Lung cancer relate
14	34.8	4.6	62944	24	ABL68262 kidney cancer rela
c 15	34.6	4.6	6741	21	AAA10595 Gene encoding a su
c 16	34.4	4.6	153	22	AAH33865 Human colon cancer
c 17	34.4	4.6	612	22	AAS22760 Human cDNA encodin
c 18	34.4	4.6	646	22	AAS22524 Human cDNA encodin
c 19	34.4	4.6	5059	20	AAH84332 Stealth virus nucl
c 20	34.4	4.6	12555	22	AAK78331 Human immune/haema
c 21	34.4	4.6	12555	23	ABK42300 Genomic sequence #
c 22	34.2	4.6	4590	22	AAH24065 Yeast AOD9604-asso
c 23	34	4.5	4268	22	ABA16693 Human nervous syst
24	34	4.5	4271	22	ABA16692 Human nervous syst
25	33.8	4.5	18189	23	ABL16916 Drosophila melanog
c 26	33.6	4.5	854	24	ABN98849 Arabidopsis thalia
c 27	33.6	4.5	1009	21	AAC43118 Arabidopsis thalia
c 28	33.6	4.5	1199	21	AAC50973 Arabidopsis thalia
c 29	33.6	4.5	1506	19	AAV16721 Nucleic acid encod
c 30	33.6	4.5	1588	22	AAH77907 Nucleotide sequenc
c 31	33.6	4.5	2472	21	AAC42997 Arabidopsis thalia
32	33.6	4.5	3413	23	ABL13645 Drosophila melanog
33	33.6	4.5	5828	23	ABL13644 Drosophila melanog
c 34	33.4	4.4	2288	21	AAA49174 cDNA encoding huma
c 35	33.4	4.4	4041	23	ABL28959 Drosophila melanog
c 36	33.4	4.4	6781	23	ABL28958 Drosophila melanog
37	33.2	4.4	10637	23	ABL09666 Drosophila melanog
38	33.2	4.4	16619	22	AAK79428 Human immune/haema
c 39	33	4.4	447	24	ABL84954 Human ovarian canc
40	33	4.4	515	22	ABA11417 Human nervous syst
41	33	4.4	1779	21	AAH44937 Ubiquitin-specific
42	33	4.4	31208	24	ABK52899 Genomic DNA encodi
43	32.8	4.4	299	22	AAS59460 Human KCC4 genomic
44	32.8	4.4	336	21	AAC36470 Human secreted pro
45	32.8	4.4	2297	23	AAS90673 DNA encoding novel

ALIGNMENTS

RESULT 1
AAA93876
ID AAA93876 standard; DNA: 10562 BP.

XX AC AAA93876;

XX DT 15-JAN-2001 (first entry)

XX DE Human beta3Gal-T5 encoding DNA.

XX KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
XX KW digestive system; beta3Gal-T5; ds.

XX OS Homo sapiens.

XX PN WO2000050608-A1.

XX PD 31-AUG-2000.

XX PF 24-FEB-2000; 2000WO-JP01070.

XX PR 25-FEB-1999; 99JP-0047571.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;

XX DR WPI; 2000-549409/50.

XX PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type I sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer

XX Claim 31; Page 103-111; 123pp; Japanese.

XX This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA; host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents a Beta3gal-T5
CC encoding DNA sequence.

XX Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 100.0%; Score 751; DB 21; Length 10562;
Best Local Similarity 100.0%; Pred. No. 2.5e-217;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGCAGACTCTAGCAGCTCAGGAGTGCACACACCTCCCGGGTTTCCTTGGTCTGGAAT 60
DB 4000 AAAGCAGACTCTAGCAGCTCAGGAGTGCACACACCTCCCGGGTTTCCTTGGTCTGGAAT 4059

QY 61 TCCATCCCTGGTTCCACCTGTTTACATCACACCTCCCTTCAAGGACGAGTCGATGC 120
DB 4060 TCCATCCCTGGTTCCACCTGTTTACATCACACCTCCCTTCAAGGACGAGTCGATGC 4119

QY 121 CACGTCTCTACGGGGTCAAGTCTACACAGTCTCTCTCCACCGAGGGCCACAGCCC 180
DB 4120 CACGTCTCTACGGGGTCAAGTCTACACAGTCTCTCTCCACCGAGGGCCACAGCCC 4179

QY 181 CTGAGACCCCTTGCAGTGTAGTCTTCTTGCATCTCTTCTGGCTCATATGCGG 240
DB 4180 CTGAGACCCCTTGCAGTGTAGTCTTCTTGCATCTCTTCTGGCTCATATGCGG 4239

QY 241 GCTTGGCCATGTCCCTTCACTCAGATCTCTCTTTCAGTCCAGGCAAGTGCATCTGA 300
DB 4240 GCTTGGCCATGTCCCTTCACTCAGATCTCTCTTTCAGTCCAGGCAAGTGCATCTGA 4299

QY 301 ACTTAACCTTCCAGACCCCTTCACTTTCAGTCTCTTAGAGAGTGGACTTCGATT 360
DB 4300 ACTTAACCTTCCAGACCCCTTCACTTTCAGTCTCTTAGAGAGTGGACTTCGATT 4359

QY 361 CTTTGTCTCTGTCCCTGTAGCTCAGTCAAGCTTAAGGCAAGTCTCTCACCTGGC 420
DB 4360 CTTTGTCTCTGTCCCTGTAGCTCAGTCAAGCTTAAGGCAAGTCTCTCACCTGGC 4419

QY 421 CTGGGAGAGTCCAGACGCTGCAGTCTGTGGGGTAGGATGCTGATGCCAGATT 480
DB 4420 CTGGGAGAGTCCAGACGCTGCAGTCTGTGGGGTAGGATGCTGATGCCAGATT 4479

QY 481 TCCGTTTAGAGAGCTTTCCTATCTCTGACGGCTTAGCTTTGTGTACTTACTTGT 540
DB 4480 TCCGTTTAGAGAGCTTTCCTATCTCTGACGGCTTAGCTTTGTGTACTTACTTGT 4539

QY 541 CCACCTTAATTCAAATGTACCCAGCAACAGCTTGTGCACAGTCTCTCTGGGTTTCAGG 600
DB 4540 CCACCTTAATTCAAATGTACCCAGCAACAGCTTGTGCACAGTCTCTCTGGGTTTCAGG 4599

QY 601 AGGATGTAAACATACCCCTTGCCTTTCAGGCACATATGCCAGAGGGGGCAGTGACC 660
DB 4600 AGGATGTAAACATACCCCTTGCCTTTCAGGCACATATGCCAGAGGGGGCAGTGACC 4659

QY 661 TAGCAGAGGGGGAGCCAGCAGATGGGATACACTCAGAGGAGCCCTGCAGCAGGCAGAG 720
DB 4660 TAGCAGAGGGGGAGCCAGCAGATGGGATACACTCAGAGGAGCCCTGCAGCAGGCAGAG 4719

QY 721 GCAGAGGAGGAGGAGTCTACAGTTCTGC 751
DB 721 GCAGAGGAGGAGGAGTCTACAGTTCTGC 751

Db 4720 GCAGAGGAGAGGAGGTCTACACGTTCTGC 4750

RESULT 2

AAS72914

XX AAS72914 standard; cDNA; 1745 BP.

AC AAS72914;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #8718.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG08727.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID No 8718; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94584 represent novel human
diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1745 BP; 306 A; 522 C; 489 G; 428 T; 0 other;

Query Match

19.4%; Score 145.8; DB 23; Length 1745;

Best Local Similarity 98.7%; Pred. No. 9.5e-34;

Matches 147; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 147 TCACACGCTTCTCTCCACGAGGGCCACAGCCCTTGGACACCCCTTGAGCTGAGTGT 206

DB 1597 TCCACAGCTTCTCTCCACGAGGGCCACAGCCCTTGGACACCCCTTGAGCTGAGTGT 1656

QY 207 TGCTTGTGACTCTTTCTGGGCTCATAGTGGGGCTTGGCCATTGTCCTCACTCCAG 266
 |||||
 Db 1657 TGCTTGTGACTCTTTCTGGGCTCATAGTGGGGCTTGGCCATTGTCCTCACTCCAG 1716
 |||||
 QY 267 ATCTCTCCCTTTCAGGTCGAGGAGTGCAT 295
 |||||
 Db 1717 ATCTCTCCCTTTCAGGTCGAGGAGTGCAT 1745
 |||||
 RESULT 3
 AAA60761/C
 ID AAA60761 standard; cDNA: 1189 BP.
 XX
 AC AAA60761;
 XX
 XX 27-OCT-2000 (first entry)
 XX
 DE Rat SENR ligand nucleotide sequence SEQ ID NO:3.
 XX
 KW SENR: sensory epithelium neuropeptide-like receptor; urotensin II;
 KW diagnosis; G protein-coupled receptor; hypertension; GPR14; hormone;
 KW kidney disease; regulator; central function; circulatory function;
 KW heart function; immune system function; digestive function;
 KW metabolic function; genital function; ss.
 XX
 OS Rattus sp.
 XX
 PN WO200032627-A1.
 XX
 XX 08-JUN-2000.
 XX
 XX 29-NOV-1999; 99WO-JP056649.
 XX
 XX 30-NOV-1998; 98JP-0338984.
 PR
 PR 04-FEB-1999; 99JP-0026848.
 PR
 PR 26-AUG-1999; 99JP-0239367.
 XX
 XX (TAKE) TAKEDA CHEM IND LTD.
 XX
 XX Mori M, Abe M, Shimomura Y, Sugo T, Kitada C;
 PI
 XX WPI: 2000-412287/35.
 XX
 XX Urotensin peptides which are ligands for sensory epithelium
 PT neuropeptide-like receptor (SENR) for diagnosis and treatment of
 PT hypertension
 XX
 PS Example 2: Page 126; 147pp: Japanese.
 XX
 CC The present invention provides peptides which are ligands for sensory
 CC epithelium neuropeptide-like receptor (SENR), and their amides, esters
 CC and salts. SENR is a G-protein coupled receptor protein (also known as
 CC GPR14), and the peptides which are ligands for it are forms of the
 CC peptide hormone urotensin II. The peptides can be used in the treatment
 CC and diagnosis of hypertension and kidney disease, and the development of
 CC drugs which are regulators of central functions, circulatory functions,
 CC heart functions, immune system functions, digestive functions, metabolic
 CC functions and genital functions. The present sequence represents a rat
 CC SENR ligand nucleotide sequence, which is used in an example from the
 CC present invention.
 XX
 XX Sequence 1189 BP; 212 A; 387 C; 309 G; 281 T; 0 other;
 SQ
 Query Match 5.8%; Score 43.6; DB 21; Length 1189;
 Best Local Similarity 51.28; Pred. No. 0.0092;
 Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
 QY 105 GGACCAAGTCGATGCCAGGTCCTTCAGGGGCTCAGAAATGCTCACCAGTCTCCTCCCA 164
 |||||
 Db 580 GGACCAAGTGGATGGCAAGCATCATGGTAGGGTCAGCAGCAGTGCCAGCAACCAAGTGTC 521
 |||||
 QY 165 CCGAGGGCCAGACGCCCTTGGAGACCCCTTGAGCTGAGTGGCTTCTCCTGACTACTCTTT 224
 |||||

Db 520 CCAGCACCCAGCAGCTTACGGTAACCCCTTGAGCGCTGGACTGTGTCC-AGAGGCCTCAGT 462
 QY 225 CTGGCCTCATAGTGGGGCTTGGCCATTGTCCTTCACTCCAGATCTCTCTTTAGGTCC 284
 |||||
 Db 461 ACGGCTGCATAGCGTTGCGCTGCTCATATATGGTCAGGGTGAAGATGCTGCGGTGCATTTGC 402
 |||||
 QY 285 AGGAAGTGCATCTTGAACCTTAACTTTCCAGACCCCTTCACTTTTCCAGTCTTTAGAG 344
 |||||
 Db 401 AGGAAGTCCAGGCTAAAGAGGACTCTGACGCCACACATCTCCAAAGTGCCAGTCTTATGTG 342
 QY 345 AGGTGG 350
 |||||
 Db 341 ACGTAG 336
 |||||
 RESULT 4
 AAZ11175/C
 ID AAZ11175 standard; cDNA: 1539 BP.
 XX
 AC AAZ11175;
 XX
 DT 04-NOV-1999 (first entry)
 XX
 DE Rat GPR14 coding sequence.
 XX
 KW GPR14: G protein-coupled receptor; ischaemic coronary artery disease;
 KW atherosclerosis; metabolic disease; CHF/myocardial dysfunction; migraine;
 KW arrhythmia; restenosis; hypertension; hypotension; pulmonary disease;
 KW fibrotic vasculopathy; cerebrovascular event; neurogenic inflammation;
 KW haematopoietic disorder; adult respiratory distress syndrome; ARDS;
 KW cancer; autoimmune disease; therapy; rat; ss.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 281..1441
 FT /*tag= a
 FT /product= GPR14
 XX
 PN WO9940192-A1.
 XX
 XX 12-AUG-1999.
 XX
 XX 27-JAN-1999; 99WO-US01634.
 PF
 PF 15-JAN-1999; 99US-0232857.
 PR
 PR 09-FEB-1998; 98US-0074075.
 PR
 PR 10-APR-1998; 98US-0058725.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM LAB PHARM.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Aiyar NV, Al-Barazanj K, Ames RS, Arnold AR, Bergsma DJ;
 PI Chambers J, Douglas SA, Foley JU, Gout B, Khandoudi N;
 PI Sarau HW, Shabon U, Willette RN;
 XX
 DR WPI: 1999-527305/44.
 DR P-PSDB: AAY32921.
 XX
 XX Human G protein-coupled receptor GPR14, useful for identifying
 PT agonists and antagonists
 XX
 PS Disclosure: Fig 2: 64pp: English.
 XX
 CC This sequence encodes the rat G protein-coupled receptor, GPR14. The
 CC invention relates to the human GPR14 protein and DNA sequences. Human
 CC GPR14, polypeptides and polynucleotides are useful in methods for
 CC treatment of, e.g. ischaemic coronary artery disease, atherosclerosis,
 CC metabolic diseases, CHF/myocardial dysfunction, arrhythmias, restenosis,
 CC hypertension and hypotension, pulmonary disease, fibrotic vasculopathies,
 CC cerebrovascular events, neurogenic inflammation/migraine, haematopoietic
 CC disorders, adult respiratory distress syndrome (ARDS), cancer, autoimmune


```
CC diseases, etc. The methods can be used to identify agonists and
XX antagonists of human and rat GPR14.
SQ Sequence 1539 BP; 291 A; 480 C; 405 G; 363 T; 0 other;

Query Match          5.8%; Score 43.6; DB 20; Length 1539;
Best Local Similarity 51.2%; Pred. No. 0.01;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 105 GGACAGTCAGATGCCACGCTTCCTACGGGGCTCAGAAATGCTACACAGTCTCTCTCCA 164
Db 854 GGACAGTCAGATGCCACGCTTCCTACGGGGCTCAGAAATGCTACACAGTCTCTCTCCA 164
QY 165 CGAGGGCCACAGCCCTGGAGACCCCTTGAGCTGAGTGTGCTTGCATACACTCTT 224
Db 794 CGAGGGCCACAGCTTACGGTAAACCTTGGAGCGCTGGATGTGTC-AGAGGCTCAGT 736
QY 225 CTGGCTCATAGTGGGGCTTGGCCATTGTCCTTCATCCAGATCTCTCTTCAGGTC 284
Db 735 ACGGCTCATAGCTTGCCTCTCATTATGCTCAGGTGAAGATGCTGGCGTCAATTGC 676
QY 285 AGGAAGTCATCTTGAACCTTAACCTTCCAGACCCCTTCAGTTTCCAGTCTCTAGAG 344
Db 675 AGGAAGTCAGCTTAAGAGGACTCTCGAGCCACATCTCCAAAGTCCAGTCTCTAGT 616

QY 345 AGGTGG 350
Db 615 ACGTAG 610

RESULT 5
AAD32925/c
ID AAD32925 standard; DNA; 1539 BP.
AC AAD32925;
XX 01-JUL-2002 (first entry)
DE Rattus norvegicus orphan GPR14 gene.
XX Rat; G-protein coupled receptor; GPCR; transgenic animal; receptor;
KW genetic therapy; pharmacological; genetic disease; neuropsychological;
KW neurological; psychotic illness; gene; nephrotropic; gynaecological;
XX psychostimulant; orphan GPR14; ds.
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT misc_feature 123..565
FT /*tag= a
FT /*note= "Sequence flanking Neo insert in targeting
FT construct"
FT CDS 281..1441
FT /*tag= b
FT /*product= "Orphan GPR14 protein"
FT 566..798
FT /*tag= c
FT /*note= "Sequence deleted in targeting construct"
FT 799..1234
FT /*tag= d
FT /*note= "Sequence flanking Neo insert in targeting
FT construct"
XX
XX WO200203793-A2.
XX
XX 17-JAN-2002.
XX
XX 10-JUL-2001; 2001WO-US21923.
XX
XX 10-JUL-2000; 2000US-217058P.
XX 10-JUL-2000; 2000US-217179P.
XX 10-JUL-2000; 2000US-217223P.
XX 10-JUL-2000; 2000US-217253P.
```

```
PR 10-JUL-2000; 2000US-217255P.
PR 10-JUL-2000; 2000US-217256P.
PR 10-JUL-2000; 2000US-217257P.
PR 11-JUL-2000; 2000US-217347P.
PR 11-JUL-2000; 2000US-217629P.
PR 12-JUL-2000; 2000US-217537P.
PR 12-JUL-2000; 2000US-218069P.
PR 12-JUL-2000; 2000US-218074P.
PR 12-JUL-2000; 2000US-218358P.
PR 27-JUL-2000; 2000US-221483P.
PR 07-AUG-2000; 2000US-223120P.
PR 07-AUG-2000; 2000US-223122P.
PR 26-OCT-2000; 2000US-243958P.
PR 15-NOV-2000; 2000US-249408P.
PR 20-NOV-2000; 2000US-252299P.
PR 16-JAN-2001; 2001US-262113P.
PR 16-JAN-2001; 2001US-262205P.
XX (DELT-) DELTAGEN INC.
XX
PI Allen KD, Brennan TJ;
XX
XX WPI; 2002-164574/21.
DR P-PSDB; AAE20600.
XX
XX Novel non-human transgenic animal, especially transgenic mice useful
PT for identifying an agent that modulates expression or function of
PT target gene, comprises disruptions in target G protein coupled receptor
PT gene .
XX
PS Disclosure; Fig 11; 101pp; English.
XX
CC The invention relates to a non-human transgenic animal having targeted
CC G-protein coupled receptor (GPCR) gene disruptions in melanocortin-3 gene
CC 5-HT-2B gene, chemokine receptor 9A gene, glucocorticoid-induced receptor
CC gene, orphan GPR10 (UHR)-11 gene, orphan GPR14 gene, orphan GPR15 gene,
CC beta chemokine receptor (E01) gene, endothelial differentiation GPCR3
CC (ED03) gene, ATP receptor P2U1 gene or adenosine 3 receptor gene. The
CC transgenic animal is useful for identifying an agent that modulates the
CC expression or function of the target gene, for identifying an agent that
CC ameliorates a phenotype associated with a disruption in the target gene.
CC A transgenic construct is useful for producing a transgenic animal,
CC preferably a transgenic mouse. The transgenic animal is useful for
CC testing the efficacy of proposed genetic and pharmacological therapies
CC for human genetic diseases, such as neurological, neuropsychological or
CC psychotic illnesses. The transgenic animal is also useful as models for
CC diseases, disorders or conditions associated with phenotypes relating to
CC a disruption in a target, and to identify pharmaceuticals, therapies,
CC drugs and interventions which may be effective in treating a disease or
CC other phenotypic characteristics of the animal. An agent which modulates
CC the expression of the target gene is useful as a therapeutic for treating
CC conditions associated with a disruption of the target gene. The present
CC sequence is rat orphan GPR14 gene.
XX
SQ Sequence 1539 BP; 291 A; 480 C; 405 G; 363 T; 0 other;

Query Match          5.8%; Score 43.6; DB 24; Length 1539;
Best Local Similarity 51.2%; Pred. No. 0.01;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps 1;

QY 105 GGACAGTCAGATGCCACGCTTCCTACGGGGCTCAGAAATGCTACACAGTCTCTCTCCA 164
Db 854 GGACAGTCAGATGCCACGCTTCCTACGGGGCTCAGAAATGCTACACAGTCTCTCTCCA 164
QY 165 CGAGGGCCACAGCCCTGGAGACCCCTTGAGCTGAGTGTGCTTGCATACACTCTT 224
Db 794 CGAGGGCCACAGCTTACGGTAAACCTTGGAGCGCTGGATGTGTC-AGAGGCTCAGT 736
QY 225 CTGGCTCATAGTGGGGCTTGGCCATTGTCCTTCATCCAGATCTCTCTTCAGGTC 284
Db 735 ACGGCTCATAGCTTGCCTCTCATTATGCTCAGGTGAAGATGCTGGCGTCAATTGC 676
QY 285 AGGAAGTCATCTTGAACCTTAACCTTCCAGACCCCTTCAGTTTCCAGTCTCTAGAG 344
```

Db	675	AGGAAGTCACGGCTAAAGAGGACTCTCGACCCACATCTCCAAAGTGCAGTCCCTAGTG	616
Qy	345	AGGTGG 350	
Db	615	ACGTAG 610	
RESULT 6			
AAA91297/c			
ID	AAA91297	standard; DNA; 3122 BP.	
XX	XX		
AC	AAA91297;		
XX	XX		
DT	19-JUN-2001	(first entry)	
XX	XX		
DE	Rat Urotensin II receptor gene.		
XX	XX		
KW	Urotensin II receptor; rat; cardiovascular disease; angina pectoris;		
KW	cardiovascular system disorder; ischaemia; hypertension; hypotension; AIDS;		
KW	myocardial infarction; stroke; congestive heart failure; shock; tumour;		
KW	impotence; nervous system disorder; trauma; immune malfunction; pain;		
KW	multiple sclerosis; myopathy; myasthenia gravis; neurodegeneration;		
KW	Parkinson's disease; amyotrophic lateral sclerosis; Alzheimer's disease;		
KW	poisoning; carbon monoxide poisoning; inborn errors of metabolism;		
KW	genetic disorder; diabetic retinal disease; retinoblastoma; depression;		
KW	psychiatric disorder; anxiety disorder; therapy; ds.		
OS	Rattus sp.		
XX	XX		
FH	Key	Location/Qualifiers	
CDS	821..1981		
FT	FT	/*tag= a	
FT	FT	/product= "Urotensin II receptor"	
XX	XX		
PN	WO200114888-A1.		
XX	XX		
PD	01-MAR-2001.		
XX	XX		
PF	03-AUG-2000; 2000WO-US21171.		
XX	XX		
PR	24-AUG-1999; 99US-0379968.		
XX	XX	(REGC) UNIV CALIFORNIA.	
PA	Civelli O, Nothacker H;		
PI			
XX	XX		
DR	WPI; 2001-226647/23.		
XX	P-PSDB; AAY97700.		
PT	Identifying urotensin II receptor modulators for treating		
PT	cardiovascular diseases, by contacting compound with the receptor which		
PT	produces specific signal in response to urotensin II, detecting		
PT	alteration in signal.		
XX	XX		
PS	Disclosure; Fig 1; 61pp; English.		
XX	XX		
CC	This sequence encodes the rat urotensin II (U II) protein. The invention		
CC	relates to a method for identifying a U II receptor agonist or		
CC	antagonist, comprises contacting an isolated U II receptor with candidate		
CC	compound(s) under conditions, where the U II receptor produces a		
CC	predetermined signal in response to U II and identifying a candidate		
CC	compound that alters production of the signal. The method is useful for		
CC	identifying an U II receptor agonist or antagonist and ligand. The		
CC	agonist or antagonist and ligands are useful for prevention or treatment		
CC	of U II receptor-associated conditions, including cardiovascular		
CC	diseases, circulatory system disorders, including ischaemia,		
CC	hypertension, hypotension, angina pectoris, myocardial infarction,		
CC	stroke, congestive heart failure, shock and impotence, disorders of the		
CC	nervous system, muscle and eye, including trauma, immune malfunction		
CC	e.g. multiple sclerosis, AIDS, myopathies, myasthenia gravis,		
CC	neurodegeneration e.g. Parkinson's disease, amyotrophic lateral		
CC	sclerosis, Alzheimer's disease, poisoning e.g. carbon monoxide poisoning		

PT New Urotensin-II Receptor (UTB-R) polypeptides and polynucleotides
PT isolated from Mus musculus, useful in gene therapy, especially for
PT treating e.g. ischemic artery disease, angina, autoimmune diseases or
XX cancers
XX
XX Claim 2: Page 31: 36pp; English.
XX
XX This DNA encodes a mouse urotensin-II receptor (UTB-R) polypeptide. The
XX UTB-R polypeptides and utb-r polynucleotides are useful in gene therapy.
XX These are useful for treating diseases, which are characterized by
XX excessive vasoconstriction, myocardial dysfunction and/or aberrant
XX fibroproliferative/inflammatory responses. In particular, they are
XX useful for treating ischemic coronary artery disease (angina and
XX myocardial infarction), atherosclerosis, metabolic diseases (e.g.
XX diabetes), hypertension, CHF/myocardial dysfunction, arrhythmias,
XX restenosis, hypertension, hypotension, pulmonary disease (e.g. COPD or
XX asthma), fibrotic vasculopathies (e.g. diabetes, SLE, AS, Reynaud's),
XX cerebrovascular events (e.g. hemorrhagic or ischemic stroke), neurogenic
XX inflammation/migraine, hematopoietic disorders, cancer, autoimmune
XX diseases (e.g. HIV-1 or -2 infection, or AIDS), gastrointestinal and
XX genitourinary disturbances (e.g. ulcers), endocrine disorders,
XX fibroproliferative disorders (e.g. psoriasis), inflammatory diseases
XX (e.g. Crohn's), benign prostatic hypertrophy, renal failure and
XX glomerulopathies, psychotic and neurological disorders (e.g. anxiety,
XX schizophrenia, manic depression, delirium, dementia, severe mental
XX retardation, Parkinson's disease or dyskinesias), infections (e.g.
XX bacterial, fungal, protozoan or viral), eating disorders (e.g. obesity,
XX anorexia or bulimia), asthma, urinary retention, osteoporosis, allergies,
XX Huntington's disease, Gilles de la Tourette's syndrome or metabolic
XX diseases (e.g. diabetes). The UTB-R polypeptide is useful as an immunogen
XX for producing antibodies that are immunospecific for UTB-R. It is also
XX useful for identifying membrane bound or soluble receptors. The utb-r
XX polynucleotides are useful for the recombinant production of UTB-R
XX polypeptides. The nucleotide sequences are also useful for chromosome
XX identification. UTB-R polypeptides, utb-r polynucleotides and antibodies
XX to UTB-R may also be used to configure screening methods for detecting
XX the effect of added compounds on the production of mRNA and polypeptide
XX in cells.
SQ Sequence 1158 BP; 193 A; 383 C; 307 G; 275 T; 0 other;
Query Match 5.2%; Score 39.4; DB 22; Length 1158;
Best Local Similarity 49.8%; Pred. No. 0.17;
Matches 126; Conservative 0; Mismatches 126; Indels 1; Gaps 1;
QY 98 CTTCAAGGACAGTGCAGATGCCACGTCCTTCACGGGGTCCAGATGCTCACCAGCTTC 157
DB 578 CCCCTACGACAGCCGCGATGGCAAGCATCATGGTAAGGTCACGACAGTGCACCAAC 519
QY 158 CTCCTCACGAGGGCCACAGCCCTCGAGACCCCTTGAGCTGAGTGTGCTTCCTTCAT 217
DB 518 CAGGTGCCAGGGCCACAGCTACGGTAACCTTGGAGCGCTGGACGGTGC-AGGG 460
QY 218 ACTCTTCTGGCCTCATAGTGGGGTGGCCATGTGCCCTCATCTCAGATCTCCCTTT 277
DB 459 CTCAGTACTGCTGATACGCTTCACTCTCATATGTCAGGGTGAAGATGCTGGCATG 400
QY 278 CAGGTCCAGGAAGTGCATCTTGAACCTTAACCTTCAGACCCCTTCAGTTTTCAGTC 337
DB 399 CATTGTCAGGAGTCCAGGCTAAGAGACTCGGACGCCACGTCGCCGAGTCCCGAGTCC 340
QY 338 CTTAGAGAGGTGG 350
DB 339 CTTAGTGACGTAG 327
RESULT 8
AAAL0594/c
ID AAAL0594 standard; DNA; 10732 BP.
XX
XX AAAL0594;
XX
XX 29-JUN-2000 (first entry)
DT

XX Gene encoding a subunit of cellulose synthase.
XX Cellulose synthase; cellulose production; increase yield; ds.
XX
XX Vigna angularis.
XX
XX JP2000060568-A.
XX
XX 29-FEB-2000.
XX
XX 26-AUG-1998; 98JP-0239998.
XX
XX 26-AUG-1998; 98JP-0239998.
XX
XX (MIZUO) MIZUNO K.
XX (OJIP) OJI PAPER CO.
XX
XX WPI; 2000-342371/30.
XX P-PSDB; AAY85179.
XX
XX A gene encoding a cellulose synthetic equipment - for the improvement
XX in the amount of cellulose synthesised in a plant body
XX
XX Claim 2: Page 14-21; 32pp; Japanese.
XX
XX This sequence represents a gene encoding a subunit of the cellulose
XX synthase complex of Vigna angularis. The invention relates to subunits of
XX cellulose synthetic equipment, that can be used to increase the amount of
XX cellulose synthesised by a plant. The proteins and genes encoding them
XX can also be used to improve the properties of the cellulose being
XX produced by a plant.
XX
XX Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;
Query Match 4.8%; Score 36.4; DB 21; Length 10732;
Best Local Similarity 26.1%; Pred. No. 4.1;
Matches 55; Conservative 70; Mismatches 84; Indels 2; Gaps 1;
QY 125 TCCTTACGGGGTCCAGATGCTCACCAGCTTCCTCTCAGCGGGCCACAGCCCTGG 184
DB 9285 BCTTVDACVTCVTSYSDCCVTCNCTCYTCCSRTBCYRAYDAYSYDANSTSTSRC 9226
QY 185 AGACCCCTGACGTGAGTGTGCTTCGATGACTCTTCTGGCCTCATAGTGGGCTT 244
DB 9225 AKYSCVYTSYSAKTSRSDSTCTCTNCYDANCCYTCYTSTNCNSTCCYTTD--YSSRST 9168
QY 245 GGCCATTGTCCTTCACTCCAGATCTCTCTTCAAGTCCAGGAGTGATCTTGAACCTT 304
DB 9167 SDYNSNSTTBTCTTSTRTCTYTSRYSDACYSNCCCTBAKCYTCSRDCSRYSTSRKTB 9108
QY 305 AACCTTCCAGACCCCTTCAGTTTCCAG 335
DB 9107 TTSTSDCVTNCNCCCVTTBDCYTNCNAK 9077
RESULT 9
AAAL08954
ID AAA08954 standard; DNA; 67212 BP.
XX
XX AAA08954;
XX
XX 01-AUG-2000 (first entry)
XX
XX WFS1 variant genomic DNA.
XX
XX WFS1; Wolfiam Syndrome; chromosome 4p; D4S500; D4S431; antidiabetic;
XX renal; opthalmic; auditory; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 8095..33109
CDS

FT FT /*tag= a
FT FT /transl_except= (pos:31801..31803, aa:Arg)
FT FT /transl_except= (pos:32267..32269, aa:His)
FT FT 8095..8326
FT FT /*tag= b
FT FT 8327..17729
FT FT /*tag= c
FT FT 17730..17812
FT FT /*tag= d
FT FT 17813..19622
FT FT /*tag= e
FT FT 19623..19767
FT FT /*tag= f
FT FT 19768..21832
FT FT /*tag= g
FT FT 21833..22003
FT FT /*tag= h
FT FT 22004..22552
FT FT /*tag= i
FT FT 22553..22633
FT FT /*tag= j
FT FT 22634..25672
FT FT /*tag= k
FT FT 25673..25821
FT FT /*tag= l
FT FT 25822..31297
FT FT /*tag= m
FT FT 31298..33109
FT FT /*tag= n
FT FT
FT FT WO200018787-A1.
FT FT
FT FT 06-APR-2000.
FT FT
FT FT 28-SEP-1999; 99WO-US22429.
FT FT
FT FT 28-SEP-1998; 98US-0102031.
FT FT
FT FT (UNIW) UNIV WASHINGTON.
FT FT (PERM/) PERMUTT M A.
FT FT (INOUE/) INOUE H.
FT FT (MUEC/) MUECKLER M.
FT FT
FT FT Permutt MA, Inoue H, Mueckler M;
FT FT
FT FT WPI; 2000-293106/25.
FT FT P-PSDB; AAY92100.
FT FT
FT FT Nucleic acids associated with, and useful for the diagnosis and
FT FT treatment of Wolfram syndrome
FT FT
FT FT Claim 7; Page 53-72; 87pp; English.
FT FT
FT FT This is the genomic DNA for a novel gene, WFS1, isolated from human
FT FT chromosome 4p, located between markers D4S500 and D4S431. Mutation of
FT FT the WFS1 gene is associated with the development of Wolfram Syndrome, a
FT FT combination of familial juvenile-onset diabetes mellitus, diabetes
FT FT insipidus, optic atrophy and deafness.
FT FT The sequence may be used as a biological marker for early diagnosis of
FT FT Wolfram syndrome and for predicting the predisposition of an individual
FT FT to the syndrome. The gene is also useful for gene replacement therapy
FT FT (i.e. to rectify the mutations and/or supplement the individuals own
FT FT production of the polypeptide) and for developing new methods and agents
FT FT (i.e. agents which bind with either the nucleic acids or the protein to
FT FT modulate it's expression and/or activity) for treating Wolfram syndrome.
FT FT It is particularly useful for generating antibodies to the protein.
FT FT
FT FT Sequence 67212 BP; 14404 A; 18346 C; 18890 G; 15571 T; 1 other;
FT FT
FT FT Query Match 4.7%; Score 35.6; DB 21; Length 67212;
FT FT Best Local Similarity 60.2%; Pred. No. 17;
FT FT Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 40 CGGGTTTCCTGTGTAATCCCATCCCGCTGTTCCACCTGTTACATCACACCTCCCC 99
Db 63407 CTGGGGTTTCGTGGGGGTGTTTCCACCCATGTTTGACGCCAGATCTTACCATCC 63466
QY 100 TTCAAGGACCAGTCAGATGCCACGTCCTTTCACGGGGC 137
Db 63467 TTCAAGGCCAGCAGCAAAATGCCACCTCTCTCCAAGAAGC 63504
RESULT 10
AAH88703
ID AAH88703 standard; DNA; 123219 BP.
AC AAH88703;
XX
XX 26-FEB-2002 (first entry)
XX
XX Human DNA sequence SEQ ID 543.
XX Single nucleotide polymorphism; SNP; biallelic marker; human;
XX central nervous system disorder; CNS; ds.
XX Homo sapiens.
XX WO200151659-A2.
XX PN
XX PD 19-JUL-2001.
XX
XX 11-JAN-2001; 2001WO-1B00116.
XX PR 13-JAN-2000; 2000US-0175854.
XX PA (GEST) GENSET.
XX
XX Chu T, Blumenfeld M, Cohen D;
XX WPI; 2001-483085/52.
XX
XX Isolated polynucleotides, useful for genotyping nucleic acids for
XX biallelic markers for the diagnosis of depression, comprises central
XX nervous system disorder related biallelic marker -
XX
XX Disclosure; Page 439-472; 519pp; English.
XX
XX The present invention relates to biallelic markers derived from human
XX genes involved in central nervous system (CNS) disorders (see
XX AAH88161-AAH88702). The markers have a single nucleotide polymorphism
XX (SNP) and are useful in determining the genetic predisposition of
XX individuals to CNS disorders, by identifying the nucleotides at a set of
XX genetic markers in a biological sample, where the markers comprise at
XX least one CNS disorder related marker. The present sequence was used
XX to illustrate the invention.
XX
XX Sequence 123219 BP; 27921 A; 32327 C; 32369 G; 30437 T; 165 other;
XX
XX Query Match 4.7%; Score 35.6; DB 23; Length 123219;
XX Best Local Similarity 60.2%; Pred. No. 23;
XX Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 40 CGGGTTTCCTGTGTAATCCCATCCCGCTGTTCCACCTGTTACATCACACCTCCCC 99
Db 118650 CTGGGGTTTCGTGGGGGTGTTTCCACCCATGTTTGACGCCAGATCTTACCATCC 118709
QY 100 TTCAAGGACCAGTCAGATGCCACGTCCTTTCACGGGGC 137
Db 118710 TTCAAGGCCAGCAGCAAAATGCCACCTCTCTCCAAGAAGC 118747
RESULT 11
AAS36579/c
ID AAS36579 standard; DNA; 3695 BP.
XX
XX AAS36579;


```
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451930/48.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
XX useful for diagnosing, treating and/or preventing disorders of the
XX cardiovascular system.
XX
XX Claim 1; SEQ ID NO 2079; 674pp; English.
XX
XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
XX the cardiovascular system antigens and their associated polynucleotides are
XX Cardiovascular system antigens and their associated polynucleotides are
XX useful in the diagnosis, treatment and prevention of various types of
XX disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX chickens or sheep. A pathological condition can be determined by
XX detecting the presence or absence of a mutation in a cardiovascular
XX system antigen polynucleotide. The treatable disorders include autoimmune
XX diseases such as rheumatoid arthritis, hyperproliferative disorders such
XX as neoplasms of the breast or liver, cardiovascular disorders such as
XX cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
XX nervous system disorders such as Alzheimer's disease, infections caused
XX by bacteria, viruses and fungi, ocular disorders such as corneal
XX infection, endocrine disorders such as premature labour and infertility,
XX gastrointestinal disorders such as Crohn's disease, renal disorders such
XX as glomerulonephritis and respiratory disorders such as asthma and
XX pleurisy. The polypeptides can also be used to aid wound healing, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, to regenerate tissues and in chemotaxis.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Query Match 4.7%; Score 35; DB 22; Length 3695;
XX Best Local Similarity 56.5%; Pred. No. 6.5;
XX Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
XX
XX QY 43 GTTTCCTGCTGGAAATCCCATCCCTGGTTCACCTGTACATCACCTCCCTC 102
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 1219 GTTTCGAATATCGGAATCTCTTACTTCTCACTGGGTAAATCTCTACCTCCCTC 1160
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 103 AAGGACCACTGAGATGCCACGCTCTTACGGGGGCTCAGAAATGCTCACGAGCTTC 157
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 1159 AAGACTCACTTCAAGGTATCACCACTAGACTACAGGATGCTTCTCTGAATCTTC 1105
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 12
XX AAS36582/c
XX ID AAS36582 standard; DNA; 3695 BP.
XX
XX AC AAS36582;
XX
XX XX
XX DT 17-DEC-2001 (first entry)
XX
XX
```

```
DE DE
XX XX
XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
XX chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
XX antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
XX cerebroprotective; neutropic; antibacterial; virucide; fungicide; cancer;
XX hyperproliferative disorder; breast; liver; autoimmune disease; neoplasm;
XX cerebrovascular disorder; nervous system disorder; bacterial infection;
XX fungal infection; viral infection; ocular disorder; endocrine disorder;
XX gastrointestinal disorder; renal disorder; respiratory disorder;
XX wound healing; skin aging; organ transplantation; tissue regeneration;
XX anti-infertility.
XX
XX Homo sapiens.
XX
XX WO200155321-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01340.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226581.
XX 22-AUG-2000; 2000US-0228668.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 06-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 08-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
```

PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232406.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246479.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451930/48.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
XX Claim 1; SEQ ID No 2082; 674pp; English.
XX
XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Query Match 4.7%; Score 35; DB 22; Length 3695;
Best Local Similarity 56.5%; Pred. No. 6.5;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 43 GTTCTTGGTCTGGAATGCCATCCCTGGTTCACCTGTATCATCACCTCCCTTC 102
Db GTTTTCTTGGTCTGGAATGCCATCCCTGGTTCACCTGTATCATCACCTCCCTTC 1160
QY 103 AGGACCACTGCAGATGCCAGCTCTTTCACGGGGCTCAGATGCTCACCAGCTTC 157
Db 1159 AAGACTCAGTTCAGGTATCACACCTAGACTACAGATGCTCTCTGAATCTTC 1105

RESULT 13
ABL66947
ID ABL66947 standard; DNA; 62944 BP.
XX
XX ABL66947;
XX
XX 15-MAY-2002 (first entry)
XX Lung cancer related gene sequence SEQ ID NO:5284.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW

PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
PI
DR WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 6599; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 62944 BP; 15145 A; 17344 C; 15900 G; 14555 T; 0 other;
Query Match 4.6%; Score 34.8; DB 24; Length 62944;
Best Local Similarity 56.8%; Pred. No. 29;
Matches 83; Conservative 0; Mismatches 62; Indels 1; Gaps 1;
QY 590 GGGGTTTCAGGAGGATGATGACATACCCCTTGCCCTTCAGGCATATGCCAGAGGG 649
DB 21951 GGGGTGGGGGTGGAACGACACGTCGCCGATTCGCCCAATCTTCTCTCAGGAAGG 22010
QY 650 GGGCAGTACCTAGGACAG-AGGGCGGGAGCCAGCAGATGGATACATCTCAGAGGCGTG 708
DB 22011 CCCAGAAACACGAGGAGGAGGAGGCGTGCGAGAACTGAATCAGTCGAGGAACCTG 22070
QY 709 CAGCAGGCAGGAGGAGGAGGAGG 734
DB 22071 AGGCAGGCAGGAGTAGTACTGGAGG 22096

RESULT 15

AAA10595/c
ID AAA10595 standard; DNA: 6741 BP.
XX
AC AAA10595;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KW Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PF 26-AUG-1998; 98JP-0239998.
XX
PR 26-AUG-1998; 98JP-0239998.
XX
PA (MIZU/) MIZUNO K.
PA (OJIP) OJI PAPER CO.
XX
DR WPI; 2000-342371/30.
DR P-PSDB; AAY85180.
XX
PT A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
XX
PS Claim 2; Page 22-31; 32pp; Japanese.
XX
CC This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
SQ Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;
Query Match 4.6%; Score 34.6; DB 21; Length 6741;
Best Local Similarity 16.3%; Pred. No. 12;
Matches 77; Conservative 181; Mismatches 210; Indels 5; Gaps 2;
QY 80 CTGTTATCATCACACCTCCCTTCAGGACCATGAGTGCAGTCCCTTCACGGGGCTC 139
DB 6115 STCCVTCYSSRRNCTTTRCTBYRACSRSTSTDSRRRCNSTTCTBTCTTTTTS 6056
QY 140 AGAATGCTCACCAGCTTCTCTCCACCGAGGGCCACACCCCTGGAGACCCCTGAGCTG 199
DB 6055 TSTTNTSTCDNS-TTBDRCDSSRGSYSSRRCTRCSTCTAKTBTBTYTDAYDARDC 5997
QY 200 AGTGTCTTCTGCTACTCTTCTGCGCTCATAGTGGGCTTGGCCATTGTCCCTTC 259
DB 5996 RCDYDASRCSTSYSDAYSSTTINSTRTBDTBNSTNSTDSRTTSRSTTYDACCVTN 5937
QY 260 ATCCAGATCTCTCTTCAGTCCAGGAGTGCATCTTGAACCTTAACCTTCCAGACCC 319
DB 5936 CSRCYTDYRCTBYRATBNCSTNSTAKRNCSTSRCTTSRNCSTTSCNSTSTNCTBDR 5877
QY 320 CCTTCAGTTCCTCAGTCTTAGAGAGGTGACTTCTGATTCCTTGTCTCTGCGCGCTG 379
DB 5876 CYRARCSTRYNSRCYSYSTYDATTBBSRGYSNSTRCDYDAYSDDYDASTDTSRSDS 5817
QY 380 TAGCCTCAGTCCAGCTTAAGCGAAGTCTCCTACCTGGCTGGGGAGAGTCCAGAC 439
DB 5816 TTTCTVSTNSTSRDNSTRTTFRCSYRCVTSYSTCTSYSDYSY---RAAKCYTSRNST 5761
QY 440 GCTGCACGTGCTGTGCGGTAGGATGCTGATGCCCAATTCCCGTTAGAGACCTTC 499

Db 5760 NSTYYDADNSTBTTTCYTTTTSINCYNSTYSYRANSTBYRRCYRARCYSCTCYT 5701
Oy 500 CCTATCCTGACGGCTCTAGCTTTGTGTGTTACTTACTTGTTCCTCCACTTTAATTC 552
Db 5700 YSYDASTCYTSRGTBCYTTBYSTBTENCRCNCYSDDSTVRANCYSSTYDATBNS 5648

Search completed: April 11, 2003, 20:22:58
Job time : 349.796 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:35:54 ; Search time 836.08 Seconds
(without alignments)
14547.424 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751

Sequence: 1 aaagcagactctagcactca.....gggaggtctacacgttctgc 751

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	237.4	31.6	523	17	AZ519305
C 2	44.6	5.9	1101	17	CNS00LXJ
C 3	42.8	5.7	1101	17	CNS0181N
C 4	40.2	5.4	997	17	CNS005TE
C 5	39	5.2	948	12	BG180681
C 6	38.8	5.2	868	17	AZ685336

7	38.8	5.2	885	17	AZ690366
8	38.2	5.1	987	17	CNS00418
C 9	38.2	5.1	1287	12	BG47752
C 10	38	5.1	580	17	AZ870115
C 11	37.8	5.0	390	17	AQ340183
C 12	37.6	5.0	1018	12	BF974400
C 13	37.6	5.0	1066	17	CNS01704
C 14	37.2	5.0	223	13	BM089196
C 15	37.2	5.0	510	17	AQ587812
C 16	37	4.9	264	10	AM481874
C 17	37	4.9	345	13	BM482333
C 18	37	4.9	382	13	BM482324
C 19	37	4.9	478	10	BE590167
C 20	37	4.9	489	13	BI682899
C 21	36.8	4.9	470	9	AA447478
C 22	36.8	4.9	703	13	BI906241
C 23	36.8	4.9	760	12	BE797465
C 24	36.8	4.9	1084	13	BM545541
C 25	36.2	4.8	403	14	B0448796
C 26	36.2	4.8	434	10	B8858536
C 27	36.2	4.8	806	17	AQ862518
C 28	36.2	4.8	943	13	BI833888
C 29	36	4.8	261	9	AT001605
C 30	36	4.8	1856	9	AL696950
C 31	35.8	4.8	674	13	BM639459
C 32	35.8	4.8	1016	17	CNS006MB
C 33	35.8	4.8	1101	17	CNS016HG
C 34	35.6	4.7	176	14	R00882
C 35	35.6	4.7	698	13	BI954345
C 36	35.6	4.7	822	12	BG762153
C 37	35.6	4.7	907	17	CNS021YL
C 38	35.6	4.7	918	14	B0671043
C 39	35.4	4.7	926	17	CNS008LM
C 40	35.2	4.7	365	10	AM438403
C 41	35.2	4.7	391	17	B06820
C 42	35.2	4.7	478	9	AJ399199
C 43	35.2	4.7	609	10	BE452204
C 44	35.2	4.7	1058	14	B0648531
C 45	35.2	4.7	1101	17	CNS00VWY

ALIGNMENTS

RESULT 1
AZ519305/C
LOCUS
DEFINITION
AZ519305
DNA sequence.
RPCI-11-91H7.TVB
RPCI-11 Homo sapiens genomic clone RPCI-11-91H7,
GSS.
ACCESSION
AZ519305
VERSION
GSS.
KEYWORDS
human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
AUTHORS
BAC end sequences of library RPCI-11
TITLE
Unpublished (1997)
JOURNAL
Other_GSSs: RPCI11-91H7.TV
COMMENT
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

TITLE Direct Submission
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information

Email: cgapbs-r@email.nill

Email: cgapbs-r@email.nill

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Borell, Oxford University Press, 1999).

BASE COUNT 291 a 199 c 165 g 230 t
ORIGIN
Query Match 5.2%; Score 38.8; DB 17; Length 885;
Best Local Similarity 54.1%; Pred. No. 7.6;
Matches 79; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 27 CCACACCCCTCCCGGGTTCTTGGTCTGGAATCCCATCCCTCCCTGTTCCACCTGTTAC 86
DB 177 CCACACCCCTCCCTGTTGAGGATGCTCCCTCCCTCCCTGTTGAGGATG 236
QY 87 ATCACACCTCCCTTCAAGSACCATGTCAGATGCCACGCTCCCTCCCTGTTGAGGATG 146
DB 237 CCTCCACCTCCCTGTTGAGGATGCTCCCTCCCTGTTGAGGATG 296
QY 147 TCACACCTCCCTCCACGGGGC 172
DB 297 AGACACCTCCCTTAACCGACG 322

RESULT 8
CNS00418 987 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BAC09C16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066537.1 GI:4942778
VERSION AL066537
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS 1 (bases 1 to 987)
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sedref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. .987
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC09C16"
/clone_lib="RPI-98"
/note="end : TET3"
BASE COUNT 238 a 162 c 17 g 177 t 393 others
ORIGIN

Query Match 5.1%; Score 38.2; DB 17; Length 987;

Best Local Similarity 23.8%; Pred. No. 12;
Matches 97; Conservative 106; Mismatches 204; Indels 0; Gaps 0;
QY 146 CTCACACCTCCCTCCACGAGGCCACAGCCCTGGAGACCCCTTGAGCTGAGTGTCT 205
DB 512 YCTTCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 571
QY 206 TTGCTTCGATACCTCTTCTGCTCATAGTGGGCTTGGCCATTGCTCCCTCACTCA 265
DB 572 TTTTTCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 631
QY 266 GATCTCTCCCTTCAGGTCAGGAGGATGTCATCTTGAACCTTCCAGACCCCTTTC 325
DB 632 YTTTCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 591
QY 326 AGTTTCACGCTTCAGAGAGGAGGAGGATGCTGATCTGCTGCTGCTGCTGCTGCT 385
DB 692 CYCTCTCCCTCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 751
QY 386 CAGTCAGGCTTAAGGCAAGCTCTCTCCTCAGCTGGGCTGGGAGAGTCCGAGGAGCT 445
DB 752 CTCCTCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 811
QY 446 CGTCCCTGTCGGGTCAGGATGCTGATGCCAGATTTCCGCTTAGAGAGGCTTTCCT 505
DB 812 TTYVCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 871
QY 506 CTGACGCTCTAGCTTGTCTGTTACTTACTTCTGCTCCTTCACTTAAATTC 552
DB 872 YTCCTCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCYCCY 918

RESULT 9
BG847752/c 1287 bp mRNA linear EST 29-MAY-2001
LOCUS 1024019A01.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BG847752
VERSION BG847752.1 GI:14228920
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 1287)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1. .1287
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

FEATURES
source
1. .1287
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda

Zip

Query Match	5.0%;	Score 37.6;	DB 12;	Length 1018;
Best Local Similarity	65.5%;	Pred. No. 18;		
Matches	55;	Conservative	0;	Mismatches 29; Indels 0; Gaps 0;
QY	642	CACAAGGGGGGAGTGACCTAGGACGAGGGGGAGGACGACGATGGGATACACTCAGAG	701	
Db	518	CAGAGCTCAGCATTTGACTGCCAGGAAGGAGGACGACTGATGACACACACTCATGG	459	
QY	702	GAGCGTCGAGCAGGAGGAGGCGACA	725	
Db	458	TGGCCTCAGACACAGCTGCTTCAGA	435	
RESULT 13				
CNS01704				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Query Match	5.0%;	Score 37.6;	DB 17;	Length 1066;
Best Local Similarity	45.3%;	Pred. No. 18;		
Matches	82;	Conservative	1;	Mismatches 98; Indels 0; Gaps 0;
QY	202	TGCTTTTGCTTCGATACCTTTCTTGGCGTCATAGTGGGCTTGGCCATTGTCCTTCAC	261	
Db	124	TGCTTTTGCTTTGNTGTTTCGCTGCTCNCCTTTTXYTNGGCGGNTGTCNNTCACTTGTC	183	
QY	262	TCAGATCTCTCCTTTCAGTGCCAGGAGTGCACTTTGAACCTTAATTTCCAGACCCGCC	321	
Db	184	TTTTCGNNTCTCTTTTGTGGGCTTGTGGTGCTKTCGSCGNTGATTGTGCTNNTTCGCT	243	
QY	322	CTTCAGTTTTCAGTGCCCTAGAGAGGTGGACTCTCTGATTCCTTTGCTCTCTGTGCCCTGTA	381	
Db	244	CGTCCTTTTGGCCNTCCCTTNCNGNNTGTTCTCTCTTTTCTTTTGTNTGTCNCTGNG	303	
QY	382	G	382	
Db	304	G	304	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:37:18 ; Search time 20.7377 Seconds
(without alignments)
11106.080 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751

Sequence: 1 aaagcagactctagcactca.....ggagggctctacacgttctgc 751

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	8.3	7218	1	US-08-232-463-14
2	43.6	5.8	1539	3	US-08-058-725B-6
3	43.6	5.8	1539	3	US-08-232-857-6
4	39.4	5.2	1158	4	US-09-587-754-1
5	32.6	4.3	2680	4	US-08-083-035-1
6	32.2	4.3	1022	4	US-08-086-105-8
7	32.2	4.3	7218	1	US-08-232-463-14
8	32	4.3	4522	5	PCT-US93-06251-22
9	31.6	4.2	1557	1	US-08-481-130-26
10	31.6	4.2	1557	1	US-08-656-984A-26
11	31.6	4.2	1557	1	US-08-485-604-26
12	31.6	4.2	1557	2	US-08-487-595-25
13	31.6	4.2	2775	1	US-08-481-130-25
14	31.6	4.2	2775	1	US-08-656-984A-25
15	31.6	4.2	2775	1	US-08-485-604-25
16	31.6	4.2	2775	2	US-08-487-595-25
17	31.6	4.2	2927	1	US-08-481-130-27
18	31.6	4.2	2927	1	US-08-656-984A-27
19	31.6	4.2	2927	2	US-08-485-604-27
20	31.6	4.2	2927	2	US-08-487-595-27
21	31	4.1	553	4	US-09-171-209-58
22	30.6	4.1	961	4	US-09-165-863-45
23	30.6	4.1	1022	4	US-09-056-105-18
24	30.6	4.1	1084	2	US-08-184-009-110
25	30.6	4.1	1084	2	US-08-458-356-110
26	30.6	4.1	1084	4	US-08-460-736-110
27	30.6	4.1	1094	2	US-08-184-009-109

Query Match

8.3%; Score 62; DB 1; Length 7218;

Sequence 109, App
Sequence 109, App
Sequence 8, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 14, Appl
Sequence 13, Appl
Sequence 14, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 14, Appl

C 28 30.6 4.1 1094 2 US-08-458-356-109
C 29 30.6 4.1 1094 4 US-08-460-736-109
C 30 30.6 4.1 1691 2 US-08-993-118-8
C 31 30.6 4.1 1691 3 US-08-845-528C-8
C 32 30.6 4.1 2419 1 US-07-807-043B-7
C 33 30.6 4.1 2419 1 US-08-299-849B-7
C 34 30.6 4.1 2419 2 US-08-142-368A-7
C 35 30.6 4.1 2419 3 US-08-967-727-7
C 36 30.6 4.1 2419 4 US-08-037-230D-7
C 37 30.6 4.1 2420 1 US-08-465-167A-23
C 38 30.6 4.1 2420 4 US-09-056-105-4
C 39 30.6 4.1 2420 4 US-08-627-820-23
C 40 30.6 4.1 2531 1 US-08-299-849B-13
C 41 30.6 4.1 2531 1 US-08-299-849B-14
C 42 30.6 4.1 2531 2 US-08-142-368A-13
C 43 30.6 4.1 2531 2 US-08-142-368A-14
C 44 30.6 4.1 2531 3 US-08-967-727-13
C 45 30.6 4.1 2531 3 US-08-967-727-14

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

Best Local Similarity	2.1%;	Pred.	No. 8.7e-09;
Matches	8;	Conservative	231; Mismatches 141; Indels 0; Gaps 0;

QY	33	CCTCCCGGGTTTCCTTGGCTGGAATCCCATCCCTGGTTCCACCTGTATACATCACCA	92
Dd	1070	YYY	1129
		: :	
QY	93	CCTCCCTTCAAGSACCAGTCAGATGCCAGTCCTTCAGGGGCCTCAGAATGCTCACCA	152
Dd	1130	YYY	1189
		: :	
QY	153	GCTTCCTTCCACGAGGGCCACAGCCCCTTGAGAGACCCCTTGAGCTGAGTGCTGTGCTCT	212
Dd	1190	YYY	1249
		: ~~~~~~ : : : : : : : : : :	
QY	213	TGCATACICTTTCTGGCCTCATAGTGGGCTTGSCCACTTCCCTTCACCTCCAGATCTCT	272
Dd	1250	YYY	1309
		: :	
QY	273	CCTTTCAGTCCAGGAAGTCATCTTGAACCTAACTTCCAGACCCCGCCTTCAGTTTTC	332
Dd	1310	YYY	1369
		: ~~~~~~ : : : : : : : : : :	
QY	333	CAGTCCTTAGAGAGTGGACTTCTGATTCCCTTGTCTCTGTGCCCTGTAGCCTCAGGTCA	392
Dd	1370	YYY	1429
		: :	
QY	393	GGCTTAAGCAAGTCTCCTCT	412
Dd	1430	YYYYYGTCACCAAAATCTTCT	1449
		: : :	

RESULT 2
US-09-058-725B-6/c
; Sequence 6, Application US/09058725B
; Patent No. 6133420
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert
; APPLICANT: Sarau, Henry
; APPLICANT: Foley, James
; APPLICANT: Chamber, Jon
; TITLE OF INVENTION: A Method of Finding Antagonist
; TITLE OF INVENTION: and Antagonist To Human and Rat GPR 14
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Smithline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,725B
; FILING DATE: April 10, 1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/789,354
; FILING DATE: 27-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T.
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GP50005-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-058-725B-6

Query Match          5.98; Score 43.6; DB 3; Length 1539;
Best Local Similarity 51.2; Patent No. 0.0015;
Matches 126; Conservative 0; Mismatches 119; Indels 1; Gaps

QY 105 GGACCAAGTCGACATGCCACGCTCCTTCACGGGCTCAGAAATGCTCACCAGCTTCCTCTCCA 164
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 854 GGACCAAGCTGGATGGCAAGCATCATGGTAGGTGACGACAGTCCCGACACACAGGTGC 195
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 165 CCGAGGGCCACAGCCCTTGGAGACCCCTTTGAGCTGAGTGCTTTTGCTCTGCATACTCTTT 224
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 794 CCAGCACCAAGCAGCTTACGGTTAACCTTGGAGCGCTGGACTGTGTGCC -AGAGGCGCTCAGT 736
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 225 CTGGGCTCATAGTGGGGCTTGGCCATGTGCCCTTCACTCCAGATCTCTCCTTTTCAGGTCC 284
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 735 ACGGTGTCATACGCTTCGCTGCTCATATATGGTCAGGGTGAAGATGCTGGCGTGCAATGTGTC 676
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 285 AGGAAGTGCATCTTGAACCTTAACCTTCCAGACCCCGCTTCCAGTCTTCAGTCTCCTTAGAG 344
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 675 AGGAAGTCCAGGCTAAGAGGACTCTGCAGCCCAATCTCCAAAGTGCAGTCTCTTAGTG 616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 AGGTGG 350
      |||||
DB 615 AGGTAG 610

RESULT 3
US-09-232-857-6/c
; Sequence 6, Application US/09232857
; Patent No. 6159700
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS, STEPHEN
; APPLICANT: WILLETTTE, ROBERT
; APPLICANT: AIYAR, NAMBI
; APPLICANT: ROMANIC, ANNE
; APPLICANT: KHANDOUDI, NASSIRAH
; APPLICANT: GOUT, BERNARD
; APPLICANT: AL-BARAZANJI, KAMAL
; APPLICANT: AMES, ROBERT S.
; APPLICANT: FOLEY, JAMES J.
; APPLICANT: SARAG, HENRY
; APPLICANT: CHAMBERS, JON K.
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DEK
; TITLE OF INVENTION: A METHOD OF FINDING AGONIST
; AND ANTAGONIST TO HUMAN AND RAT GPR14
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/232,857
; FILING DATE: 15-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/789,354
; FILING DATE: 27-JAN-1997
; APPLICATION NUMBER: 60/074,075

```

Db	578	CCCCTACGGACCCAGCGGATGGCAAGCATCATGGGTAAAGTTCAGCAGCAGTGCACGACAC	51
Qy	158	CTCTCCACCGAGGCGCCACAGCCCTGGAGACCCCTTGAGCTGAGTGCCTTTGCTCTTGCAT	217
Db	518	CAGGTGCCAGCGCCAGACGCTTACGGTAAACCTTGGAGCGCTGGACGGTGCC-AGGGG	460
Qy	218	ACTCTTTCTGGCCTCATAGTGGGCGCTTGGCCATTGGTCCCTTCACTCCAGATCTCTCTTT	277
Db	459	CCTCAGTACTGCTGCATAGCGTTTCACTGCTCATTTATGGTCAGGCTGAAGATGCTGCGATG	400
Qy	278	CAGGTCCAGGAAGTGCATCTTTGAACCTTTAACTTTCCAGACCCCTTTCAGTTTCCAGTC	337
Db	399	CATTGTCAGGAAGTCCAGGCTTAAAGAGAACTCGGCAGCCCACTGCCGAAAGTGCCAGTC	340
Qy	338	CTTAGAGAGGTGG	350
Db	339	CTTAGTGACGTAG	327
<p>RESULT 5</p> <p>US-09-063-035-1/c</p> <p>Sequence 1, Application US/09063035</p> <p>Patent No. 6160091</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: PEUKERT, Karen; HAENEL, Frank; and EILERS,</p> <p>APPLICANT: MARTIN</p> <p>TITLE OF INVENTION: Myc-binding zinc finger proteins,</p> <p>TITLE OF INVENTION: their preparation and their use</p> <p>NUMBER OF SEQUENCES: 2</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Keil & Weinkauf</p> <p>STREET: 1101 Connecticut Avenue</p> <p>CITY: Washington</p> <p>STATE: D.C.</p> <p>COUNTRY: USA</p> <p>ZIP: 20036</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage</p> <p>COMPUTER: IBM AT-compatible, 80486 processor</p> <p>OPERATING SYSTEM: MS-DOS version 6.1</p> <p>SOFTWARE: WordPerfect version 8.0</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/063,035</p> <p>FILING DATE: 21-APR-1998</p> <p>CLASSIFICATION: 514</p> <p>INFORMATION FOR SEQ ID NO: 1:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 2680 base pairs</p> <p>TYPE: Nucleic acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: cDNA to mRNA</p> <p>HYPOTHETICAL: NO</p> <p>ANTI-SENSE: NO</p> <p>FEATURE:</p> <p>NAME/KEY: 5'UTR</p> <p>LOCATION: 1..159</p> <p>FEATURE:</p> <p>NAME/KEY: CDS</p> <p>LOCATION: 160..2571</p> <p>FEATURE:</p> <p>NAME/KEY: CDS</p> <p>LOCATION: 2572..2680</p> <p>US-09-063-035-1</p>			
<p>Query Match 4.3%; Score 32.6; DB 4; Length 2680;</p> <p>Best Local Similarity 52.6%; Pred. No. 4.3;</p> <p>Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps</p>			
Qy	267	ATCTCTCTTTTCAGGTCCAGGAAGTGCATCTTGAACCTTAACTTCCAGACCCCTTCA	326
Db	338	ACCAGTCTTCTGCTGCCAGAGAGCATCTTGAAGTACTCTGCTGCAGCCCGCCAGCACT	279


```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-22

Query Match          4.3%; Score 32; DB 5; Length 4522;
Best Local Similarity 50.7%; Pred. No. 8.4;
Matches 77; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 570 CAGCTTGTCCAGAGTCTCTCTGCGGTTTCAGAGGAGTGTAGACATACCCCTTGCCTTC 629
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3910 CAGGCTCTCCACCCCTGCGGAGAGTATAGGAGGACTTAGGAGACCCGAGAAAACGC 3851
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 630 AGCAGCTATGGCCAGGAAGGGGGGCGAGTGCCTTAGCAGAGGGGGGAGCCAGCAGATGGG 689
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3850 AGACAAAGGGGAAGAGAGAGCAGAGAGGTGCGTGGCGGATGGAGCAGGCGAGATGA 3791
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 690 ATACACTCAGAGAGCGCTGCAGCAGCGAGG 721
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3790 CACCATCAGGAGCGCAGCCACCCCTGCGGAGGGG 3759
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-08-481-130-26/c
; Sequence 26, Application US/08481130
; Patent No. 5702917
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Kilgannon, Patrick D.
; TITLE OF INVENTION: ICAM-4 Materials and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.130
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,689
; FILING DATE: 27-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,724
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/894,061
; FILING DATE: 05-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,266
; FILING DATE: 22-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,852
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,295
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, JR. JOSEPH A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32713
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1557 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-481-130-26

Query Match          4.2%; Score 31.6; DB 1; Length 1557;
Best Local Similarity 56.9%; Pred. No. 6.6;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 129 TCACGGGGCTCAGAAATGCTCACCAGCTTCTCTCCACGAGGCCACAGCCCTCGAGAC 188
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 322 TCCTCCGGCTCGGAAATGCTCACCAGCTGCGGGCCCAACCAACGCAACCCCTCTGGTCC 263
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 189 CCCTTGAGCTGATGCTGTTGCTTGTGCATACACTCTTTTCTGTGCC 230
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 CGTTTCGGCGCAGCAGGAGTCTCCAGGCCACCGCGGTCCGGCC 221
    ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-08-656-984A-26/c
; Sequence 26, Application US/08656984A
; Patent No. 5753502
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Kilgannon, Patrick D.
; TITLE OF INVENTION: ICAM-4 Materials and Methods
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/656.984A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,689
; FILING DATE: 27-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,724
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/894,061
; FILING DATE: 05-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,266
; FILING DATE: 22-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,852
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,295
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, JR. JOSEPH A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33321
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1557 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-656-984A-26

Query Match 4.2%: Score 31.6; DB 1; Length 1557;
Best Local Similarity 56.9%; Pred. No. 6.6;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 129 TCACGGGGCTCAGAATGCTCACCAGCTTCCTCTCCACGGAGGGCCACAGCCCTGTGGAGAC 188
DB 322 TCTCGGGCTCGGGAATGTCACAGCTGCGCGGCCAACCAACGCAACCCCTCTGGGTCC 263
QY 189 CCCTTGAGCTGAGTCTTGTGCTTGCATCTCTTCTGGCC 230
DB 262 CGTTTCGGCGCAGCAGGTCTCCAGGCCACCGCGCTCCGGCC 221

RESULT 11
US-08-485-604-26/c
Sequence 26, Application US/08485604
Patent No. 5773293
GENERAL INFORMATION:
APPLICANT: WP, W. Michael
APPLICANT: Kilgannon, Patrick D.
TITLE OF INVENTION: ICAM-4 Materials and Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,604
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,689
FILING DATE: 27-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,724
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/894,061
FILING DATE: 05-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,266
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,852
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,295
FILING DATE: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, JR. JOSEPH A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32715
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1557 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-485-604-26

Query Match 4.2%: Score 31.6; DB 1; Length 1557;
Best Local Similarity 56.9%; Pred. No. 6.6;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 129 TCACGGGGCTCAGAATGCTCACCAGCTTCCTCTCCACGGAGGGCCACAGCCCTGTGGAGAC 188
DB 322 TCTCGGGCTCGGGAATGTCACAGCTGCGCGGCCAACCAACGCAACCCCTCTGGGTCC 263
QY 189 CCCTTGAGCTGAGTCTTGTGCTTGCATCTCTTCTGGCC 230
DB 262 CGTTTCGGCGCAGCAGGTCTCCAGGCCACCGCGCTCCGGCC 221

RESULT 12
US-08-487-595-26/c
Sequence 26, Application US/08487595
Patent No. 5852170
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Kilgannon, Patrick D.
TITLE OF INVENTION: ICAM-4 Materials and Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,595
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,689
FILING DATE: 27-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,724
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/894,061
FILING DATE: 05-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,266
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,852
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,295
FILING DATE: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, JR. JOSEPH A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32714
TELECOMMUNICATION INFORMATION:

FILING DATE: 16-MAY-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/485,604
 FILING DATE: 07-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: WILLIAMS, JR. JOSEPH A.

REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33321
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 2775 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-656-984A-25

Query Match 4.2%; Score 31.6; DB 1; Length 2775;
Best Local Similarity 56.9%; Pred. No. 8.8;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 129 TCACGGGGCTCAGATGCTCACAGCTTCCTCTCCACCGAGGGCCACAGCCCTGGAGAC 188
DB 170 TCTCCGGCTCGGATGTCCACACGCTCGCGGCCAACCAACGCAACCCCTCTGGGTCC 111
QY 189 CCCTTGAGCTGAGTGTCTTGCTGCTTCCTGCTATCTCTTTCTGGCC 230
DB 110 CTTTGGCGCAGCGAGGTCTCCAGGCCACCGCGCTCCGGCC 69

RESULT 15
US-08-485-604-25/C
Sequence 25, Application US/08485604
Patent No. 5773293
GENERAL INFORMATION:
APPLICANT: WP, W. Michael
APPLICANT: Kilgannon, Patrick D.
TITLE OF INVENTION: ICAM-4 Materials and Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,604
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,689
FILING DATE: 27-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,724
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/894,061
FILING DATE: 05-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/009,266
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,852
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,295
FILING DATE: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, JR. JOSEPH A.

REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32715
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 2775 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-485-604-25

Query Match 4.2%; Score 31.6; DB 1; Length 2775;
Best Local Similarity 56.9%; Pred. No. 8.8;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 129 TCACGGGGCTCAGATGCTCACAGCTTCCTCTCCACCGAGGGCCACAGCCCTGGAGAC 188
DB 170 TCTCCGGCTCGGATGTCCACACGCTCGCGGCCAACCAACGCAACCCCTCTGGGTCC 111
QY 189 CCCTTGAGCTGAGTGTCTTGCTGCTTCCTTGCATCTCTTTCTGGCC 230
DB 110 CTTTGGCGCAGCGAGGTCTCCAGGCCACCGCGCTCCGGCC 69

Search completed: April 11, 2003, 22:48:18
Job time : 43.7377 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 21:30:39 ; Search time 48.0484 Seconds
(without alignments)
13710.183 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751
Sequence: 1 aaagcagactctagcactca.....gggaggtctacacgttctgc 751

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:**
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:**
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:**
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:**
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:**
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:**
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:**
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:**
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:**
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:**
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:**
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:**
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:**
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38.6	5.1	671	9	US-10-184-644-346
C 2	36.8	4.9	655	9	US-10-184-644-418
C 3	35.4	4.7	439	10	US-09-983-965-5682
C 4	35.2	4.7	632	9	US-10-184-644-166
C 5	35	4.7	3695	9	US-10-091-504-2079
C 6	35	4.7	3695	9	US-10-091-504-2082
C 7	35	4.7	3695	9	US-09-764-869-2079
C 8	35	4.7	3695	10	US-09-764-869-2082
C 9	34.8	4.6	62944	10	US-09-954-456-2257
C 10	34.4	4.6	12555	9	US-10-092-154-1187
C 11	34.4	4.6	12555	10	US-09-764-847-1187
C 12	34.2	4.6	395	9	US-09-894-844-45
C 13	33.6	4.5	351	9	US-10-184-644-80
C 14	33.6	4.5	854	10	US-09-770-445-617
C 15	33.6	4.5	1060	10	US-09-833-381-1503
C 16	33.6	4.5	2472	9	US-09-938-842A-677
C 17	33.4	4.4	3991	9	US-10-108-605-348
C 18	33.2	4.4	2000	9	US-09-938-842A-4149
C 19	33	4.4	447	10	US-09-867-701-7932

20	33	4.4	31208	10	US-09-852-067-3	Sequence 3, Appli
21	32.8	4.4	239	9	US-09-835-9758-82	Sequence 82, Appl
C 22	32.8	4.4	7275	9	US-09-764-872-864	Sequence 864, App
C 23	32.6	4.3	1533	10	US-09-815-242-7755	Sequence 7755, Ap
C 24	32.6	4.3	73467	9	US-10-237-859-3	Sequence 3, Appli
C 25	32.4	4.3	948	10	US-09-770-445-343	Sequence 343, App
C 26	32	4.3	1997	10	US-09-925-301-555	Sequence 555, App
C 27	31.8	4.2	392	9	US-10-184-644-160	Sequence 160, App
C 28	31.8	4.2	470	10	US-09-864-761-1732	Sequence 1732, Ap
C 29	31.8	4.2	2448	10	US-09-764-877-3272	Sequence 3272, Ap
C 30	31.8	4.2	3162	9	US-09-738-626-2352	Sequence 2352, Ap
C 31	31.8	4.2	3174	10	US-09-895-652-19	Sequence 19, Appl
C 32	31.6	4.2	133893	9	US-10-161-510-1	Sequence 1, Appli
C 33	31.6	4.2	393	10	US-09-983-965-3870	Sequence 3870, Ap
C 34	31.6	4.2	31814	10	US-09-817-182-3	Sequence 3, Appli
C 35	31.4	4.2	302250	10	US-09-962-832-154	Sequence 154, App
C 36	31.2	4.2	323	10	US-09-880-107-1024	Sequence 1024, Ap
C 37	31.2	4.2	545	9	US-10-184-644-58	Sequence 58, Appl
C 38	31.2	4.2	564	10	US-09-864-761-8606	Sequence 8606, Ap
C 39	31.2	4.2	653	9	US-10-184-644-402	Sequence 402, App
C 40	31.2	4.2	726	10	US-09-864-761-19707	Sequence 19707, A
C 41	31.2	4.2	1446	9	US-09-938-842A-1812	Sequence 1812, Ap
C 42	31.2	4.2	1951	10	US-09-864-761-2926	Sequence 2926, Ap
C 43	31.2	4.2	22609	10	US-09-764-877-2146	Sequence 2146, Ap
C 44	31.2	4.2	32193	10	US-09-764-877-2147	Sequence 2147, Ap
C 45	31	4.1	553	9	US-10-228-794-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1

US-10-184-644-346/c
; Sequence 346, Application US/10184644
; Publication No. US20030044930A1

GENERAL INFORMATION:

- ; APPLICANT: Baker, Kevin P.
- ; APPLICANT: Chen, Jian
- ; APPLICANT: Desnoyers, Luc
- ; APPLICANT: Goddard, Audrey
- ; APPLICANT: Godowski, Paul J.
- ; APPLICANT: Gurney, Austin L.
- ; APPLICANT: Pan, James
- ; APPLICANT: Smith, Victoria
- ; APPLICANT: Watanabe, Colin K.
- ; APPLICANT: Wood, William I.
- ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644

; PRIOR APPLICATION DATE: 2002-06-28

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 346

; LENGTH: 671

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-184-644-346

Query Match 5.1%; Score 38.6; DB 9; Length 671;

Best Local Similarity 6.6%; Pred. No. 0.022;

Matches 24; Conservative 143; Mismatches 199; Indels 0; Gaps 0;

QY	47	CGTTGGTCTGGAAATCCCATCCCTGGTTCACCTGTTACATCACACCTCCCTCCAGG	106
DB	495	YMBH.SHB.M.T.M..SD.M.M..BS..MM.BMYHS.C.SVMYAY.BMB.MT..M....	436
QY	107	ACCAGTGCAGATCCCGAGTCCTTCACGGGGCTCAGATGCTCACCAGTCTCTCCACC	166
DB	435	YBYM.M.C.MYT..ASSS..MMTSMNT.Y...T...SHSS..CYCYMYCYMMB..HH.	376
QY	167	GAGGGCCACAGCCCTGGAGACCCCTTGAGTGGCTTTTGCTTGTGCATCTCTTCT	226

Db 375 Y...HSSCSCV.T.CYHTY.YYY.M.M...Y.YY.....YYYYT...YYT.HYMW 316
QY 227 GGCCTCATAGTGGGCTTGGCCATGTCCTCACTCCAGATCTCTCTTTCAGGTCAC 286
Db 315 .S.YHB.HSHSSSSS..Y..M.MYC.M...M.T.MYC..MMSBHSBSSSSSSSS 256
QY 287 GAAGTCATCTGACCTTAACCTTCCAGACCCCCCTTCAGTTTCCAGTCCTTAGAG 346
Db 255 STSYKT.B...MTCSHSMHSHSTS.TMMMYCC...CY.TYETMM...A.H.HSAM.S 196
QY 347 GTGGACTTCTGATCTCTGTCCTGTCCTGAGCTCAGGTCAGGCTTAAGGCAAG 406
Db 195 .SS.SN...S.SBST.H.HSTWYMSBSKM.T.AMYM.CSNHSSMHSBHS.KYHST 136
QY 407 TCTCCT 412
Db 135 TATABT 130
RESULT 2
US-10-184-644-418/c
; Sequence 418, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Collin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 418
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-418

Query Match 4.9%; Score 36.8; DB 9; Length 655;
Best Local Similarity 9.0%; Pred. No. 0.081; Mismatches 121; Indels 0; Gaps 0;
Matches 36; Conservative 121; Mismatches 121; Indels 0; Gaps 0;
QY 31 CACCTCCCGGGTTTCCTTGGTGGATTCACATCCCTCCGTTCCACCTGTATACATCA 90
Db 574 CNY.S.TSSCSASHG....S.HB..CMNMH..S.BA..TSN..MTN.S...S.S.S. 515
QY 91 CACCTCCCTTCAAGACAGTCAGATGCCAGTCCTTCACGGGGCTCAGAATGCTCAC 150
Db 514 SK..T.MHA...AAH.K.CY.M.BHNYD.Y.TS...T.ST..CY.AWD..TTRTY.DH 455
QY 151 CAGCTTCCTCTCCACCGAGGCCACGCTCCCTGGAGACCCCTTGAGCTGAGTGTGTC 210
Db 454 TARCNS.TTB.Y.STNG...R.HMW.SCB.TTB.M..H.CDCNGRR.WM.YN.A.AKSMM 395
QY 211 CTGTCATCTCTTTCGGCTCATAGTGGGCTTGGCCATGTCCTTCCTCACTCCAGATCT 270
Db 394 .CTM.B.TS.H.Y.CMM.AYSSM.SGBB.BB....B.KW..D.N.H.DMD.N.Y.D 335
QY 271 CTCTTTCAGGTCAGGAGTGCATCTCACTTAACCTTTCAGACCCCTTCAGTTT 330
Db 334 CYM.CH..ASSM.CCAT.KS...M.DYDD.C..DNBB..N..AMNBH.MCYTSSANH.B 275

QY 331 TCCAGTCTTAGAGAGGTGGACTTCTGATTCTCTGTCTGTGCTGCTAGCCTCAGGT 390
Db 274 AC...SS.BM.YBSTSSNS.ASNKM.BRASS.B.DA.KD..Y...TAC.S.SASSS.S. 215
QY 391 CAGGCTTAAAGCAAGTCTCTCACTCACTGCGCTGGGAG 428
Db 214 .ACGENHA.MAC.M.BB.N.S.GHARTMGMBSS.BHS 177
RESULT 3
US-09-983-965-5682
; Sequence 5682, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION A
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5682
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 55-LIB34-050-Q1-E1-F4
US-09-983-965-5682
Query Match 4.7%; Score 35.4; DB 10; Length 439;
Best Local Similarity 57.8%; Pred. No. 0.2;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 299 GAACCTTAACCTTCCAGACCCCTTCCAGTCTTCCAGTCTTAGAGAGTGGACTTCGA 358
Db 315 GCATTTACGTTTGCAGACTCAAAATCCAGCTCTCCAGTCTCCGTCGATGGGGTCTCTGG 374
QY 359 TTCTTTGTCTGTGTCCTGTAGCTCAGTCTCAGGTCAGGCTTAAGGCAAGT 407
Db 375 TCCACCGTTGGTCCATCTTAGAGACTCAGGCCACCTTTCGGCCAGAT 423
RESULT 4
US-10-184-644-166/c
; Sequence 166, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Collin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 166

```
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-166

Query Match          4.7%; Score 35.2; DB 9; Length 632;
Best Local Similarity 8.5%; Pred. No. 0.26;
Matches 39; Conservative 147; Mismatches 261; Indels 1; Gaps 1;

QY 41 GGTTCCTTGCTGGAATCCCATCCCTGCTGCTCCACCTGTTACATCACACCTCCCTC 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 515 DNSH.TT.SSGH...R.MB...TM.B.SSSAYM..TBT.SYSB.A..BCHBN...HCAM. 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 101 TCAGGACGAGTCAGATCCAGCTCCTTCACGGGGCTCAGAGTCTCACCAGCTTCCTC 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 YCHYS.BCC..SS.BR...HW.YDSTCCTBAKC.S.C.HM..NBEM.DGA.A..D..M.A 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 161 TCACCGAGGGCCAGACCCCTGGAGACCCCTTGAGCTGAGTGCCTTGCTTCGTGATCT 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 NSY.C.C.SWSCNSNCT...H.SY.YB.YSBB.DBYI.ST...DTTS...SCRY.HDCN 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 221 CTTCCTGCTCATAGTGGGGCTGGCCATTGTCCTTCCTCCTCAGATCTCTCTTCAG 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 .T.BYHN...CDYRTBCCH.BN...BC...HBMYB.M.C.....SSMN..BD.SHY.YR 277
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 281 GTCAGGAGTGCATCTTGAAGTTAACTTTCCAGACCCCTTCAGTTTCCAGTCTCCT 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 TH.T.CNNYSY.M..YKBA.W.B.G.Y..YBTRND.BNS.HKCNBM...HC...YCHYT 217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 341 AGAGAGTGGAGTCTTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 .BCHYR.D...DB..A.SCCB.Y.S.S.S.HBYN.M.SA..CH...DR.Y..B..TAAN 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 401 GCAAGGTCTCTCCTCAGCTGGGGAGAGTCCAGAGGCTGCGCTGCTGCGGGT 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 .SN.YC.HDNTFTSCMAYI.BS.TYN.MM.SYSY.ASY..YTYAYNSYCSHB.S.T. 97
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 461 AGGATGCTGATGCCAGATTCCTCGGT 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96 .CH.TSSBRT.NH.C...HAK.S.ATTT 70
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-10-091-504-2079/c
; Sequence 2079, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2079
; LENGTH: 3695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-2079

Query Match          4.7%; Score 35; DB 9; Length 3695;
Best Local Similarity 56.5%; Pred. No. 0.58;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 43 GTTTCCTTGCTGGAATCCCATCCCTGCTGCTCCACCTGTTACATCACACCTCCCTTC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 GTTTCCAATATCTGGAATTCCTTCTACTTCTCACTGCTAATTCCTACCCGTCCTC 1160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 103 AAGACCACTGAGTGCACGCTCTTCACGGGGCTCAGAGTCTCACCAGCTTC 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 AAGACTCAGTTCAGGTATCACCACCTAGACTACAGGATGCTTCTCTGAATCTTC 1105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-10-091-504-2082/c
; Sequence 2082, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2082
; LENGTH: 3695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-2082

Query Match          4.7%; Score 35; DB 9; Length 3695;
Best Local Similarity 56.5%; Pred. No. 0.58;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 43 GTTTCCTTGCTGGAATTCCTCCATCCCTGCTGCTCCACCTGTTACATCACACCTCCCTTC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 GTTTCCAATATCTGGAATTCCTTCTACTTCTCACTGCTAATTCCTACCCGTCCTC 1160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 103 AAGACCACTGAGTGCACGCTCTTCACGGGGCTCAGAGTCTCACCAGCTTC 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 AAGACTCAGTTCAGGTATCACCACCTAGACTACAGGATGCTTCTCTGAATCTTC 1105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-764-869-2079/c
; Sequence 2079, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2079
; LENGTH: 3695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-2079

Query Match          4.7%; Score 35; DB 10; Length 3695;
Best Local Similarity 56.5%; Pred. No. 0.58;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 43 GTTTCCTTGCTGGAATTCCTCCATCCCTGCTGCTCCACCTGTTACATCACACCTCCCTTC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 GTTTCCAATATCTGGAATTCCTTCTACTTCTCACTGCTAATTCCTACCCGTCCTC 1160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 103 AAGACCACTGAGTGCACGCTCTTCACGGGGCTCAGAGTCTCACCAGCTTC 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 AAGACTCAGTTCAGGTATCACCACCTAGACTACAGGATGCTTCTCTGAATCTTC 1105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-09-764-869-2082/c
; Sequence 2082, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
```

Db 22011 CCCAGAAAACCCAGGAGGAAGGGCGTGCACAATCACTAGTCAGGAGAACCCTG 22070

Qy 709 CAGCAGGCCAGGAGGAGAAAGGG 734
| | | | |
| | | | |

Db 22071 AGGCAGGCCAGGTACTTGAGGG 22096
| | | | |
| | | | |

RESULT 10
US-10-092-154-1187/c
; Sequence 1187, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1187
; LENGTH: 12555
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1187

Query Match 4.6%; Score 34.4; DB 9; Length 12555;
Best Local Similarity 59.0%; Pred. No. 1.4;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0

Qy 637 ATGCCCAAAGGGGGCAGTACCTAGGAGGCGGGGGAGCAGCATGGATACACT 696
| | | | |
| | | | |

Db 4879 AGGCCAGGATGGGGCAATCCCTGGGGCAGGTTGAGCAGAGAGGTGGGCTGGGTG 4820
| | | | |
| | | | |

Qy 697 CAGAGGAGCCTGCAGCAGGCGAGCGGAGAGAGGGAG 736
| | | | |
| | | | |

Db 4819 GAGACTAGGTGAGGCAGGAGCAATGGCAGAGGAAG 4780
| | | | |
| | | | |

RESULT 11
US-09-764-847-1187/c
; Sequence 1187, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1187
; LENGTH: 12555
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1187

Query Match 4.6%; Score 34.4; DB 10; Length 12555;
Best Local Similarity 59.0%; Pred. No. 1.4;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0

Qy 637 ATGCCCAAAGGGGGCAGTACCTAGGAGGCGGGGGAGCAGCATGGATACACT 696
| | | | |
| | | | |

Db 4879 AGGCCAGGATGGGGCAATCCCTGGGGCAGGTTGAGCAGAGAGGTGGGCTGGGTG 4820
| | | | |
| | | | |

Qy 697 CAGAGGAGCCTGCAGCAGGCGAGCGGAGAGAGGGAG 736
| | | | |
| | | | |

Db 4819 GAGACTAGGTGAGGCAGGAGCAATGGCAGAGGAAG 4780
| | | | |
| | | | |

RESULT 12
US-09-894-844-45/c

Sequence 45, Application US/09894844
 Patent No. US20020176873A1
 GENERAL INFORMATION:
 APPLICANT: Behr, Marcel
 APPLICANT: Small, Peter
 APPLICANT: Schoolnik, Gary
 APPLICANT: Wilson, Michael A.
 TITLE OF INVENTION: Molecular Differences Between Species of
 FILE REFERENCE: STAN102CON the M. Tuberculosis Complex
 CURRENT APPLICATION NUMBER: US/09/894,844
 CURRENT FILING DATE: 2001-06-27
 PRIOR APPLICATION NUMBER: 09/318,191
 PRIOR FILING DATE: 1999-05-25
 PRIOR APPLICATION NUMBER: 60/097,936
 PRIOR FILING DATE: 1998-08-25
 NUMBER OF SEQ ID NOS: 137
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 45
 LENGTH: 395
 TYPE: DNA
 ORGANISM: Mycobacteria tuberculosis
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(395)
 OTHER INFORMATION: n = A,T,C or G
 NAME/KEY: misc_feature
 LOCATION: 27, 44, 104, 119, 180, 224, 237, 245, 254, 301, 327, 370,
 LOCATION: 365, 393
 OTHER INFORMATION: n = A,T,C or G
 US-09-894-844-45

Query Match 4.6%; Score 34.2; DB 9; Length 395;
 Best Local Similarity 18.3%; Pred. No. 0.46;
 Matches 58; Conservative 125; Mismatches 134; Indels 0; Gaps 0;
 QY 33 CCTCCCGGGTTCTTGGCTGGAATCCCATCCCTGGTTCACCTGTATCATCACA 92
 Db 317 CBTYYHCDTHCTCCGNHVTBTBBRRTHTSATATCBKBYCYCTBTBBBHTCAYCACM 258
 QY 93 CPTCCCTTCAAGGACGATGCAGATGCACGCTCCTTCACGGGGCTCAGATGCTCACC 152
 Db 257 CSRNMSHTWCSBNHYHCCSSNCDSTACCYWTNSBBBACCTWYWTBTBCCBMBTBA 198
 QY 153 GCTTCTCTCCACCGAGGCGCACGCCCTGTGAGACCCCTGTAGCTGAGCTTGCTT 212
 Db 197 ATCWYTSBRTTBBCASNCYHBCWTTTTCABBBYCYCTTWTSTKATBRGCTCWDSC 138
 QY 213 TGCATACCTTCTTGGCCCTCATAGTGGGCTTGCCCATTTGCCCTTCACTCCAGATCTCT 272
 Db 137 TRTTTWTBTAHTSRHTWCTTTBTRHSGDYCNKCHWGHCBASBHTBWTSTRTATS 78
 QY 273 CTTTCAGTCCAGGAGTGCATCTTGAACCTTAACCTTCCAGACCCCTTTCAGTTTTC 332
 Db 77 BSTACKAYSSCTWYSTABTCYBYMKYTYTCDNCBYHCVSYMYATBSBHNCTCBHGC 18
 QY 333 CAGTCCTTAGAGAGGTG 349
 Db 17 YCSADYTBCAVTSKB 1

RESULT 13
 US-10-184-644-80/c
 Sequence 80, Application US/10184644
 Publication No. US2003004930A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Chen, Jian
 APPLICANT: Desnoyers, Luc
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Pan, James

APPLICANT: Smith, Victoria
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3430R1C227 ACIDS ENCODING THE SAME
 CURRENT APPLICATION NUMBER: US/10/184,644
 CURRENT FILING DATE: 2002-06-28
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 612
 SEQ ID NO 80
 LENGTH: 351
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-10-184-644-80
 Query Match 4.5%; Score 33.6; DB 9; Length 351;
 Best Local Similarity 11.7%; Pred. No. 0.68;
 Matches 35; Conservative 95; Mismatches 170; Indels 0; Gaps 0;
 QY 80 CTGTTACATCACACCTCCCTTCAAGGACGATGCAGATGCCACGCTCCTTACGGGGCTC 139
 Db 305 CH.TMSAMNGAYCYAC.BCSYKH..G.H.S...RB.H.HAD.M..TNY.B.TYSCBYT. 246
 QY 140 AGAATGCTCACCAGCTTCTCTCCACGAGGCGCCACGAGCCCTGGAGACCCCTTGAGGTG 199
 Db 245 .B.ATCH.M.M.TDCB.Y...BTYWGAMB.GSCSBCDGMG.BYKDA..TMYCT.NND.NK 186
 QY 200 AGTGTCTTGTGCTGATCTCTTCTGGCTCATAGTGGGCTTGGCCCATTTGCCATTTCCAGACCCC 259
 Db 185 .TYSSTTCSY.YBHB.S.S.TBCRT.NHSGGSW..C..SBGDBAYHCGCM...CSSGT 126
 QY 260 ACTCCAGATCTCTCTTCCAGGTCAGGAGTGCATCTTGAACCTTAACCTTCCAGACCCC 319
 Db 125 ABT.TBCTSS.TRB.TT.YAC.ABSMC.B..SH.ASGNWIYNY..R.G...T..TCYBS 66
 QY 320 CCCTTCAGTTTCCAGTCTCTTACGATGAGAGGTGAGCTTCTGATCTTGTCTGTGCTGCTG 379
 Db 65 HKB..NYMGK.B.Y...CM.M.GA...S.CBSS.MT.R.WNSTTTS.BT.B..Y.SY.G 6

RESULT 14
 US-09-770-445-617/c
 Sequence 617, Application US/09770445
 Patent No. US2002023281A1
 GENERAL INFORMATION:
 APPLICANT: Goriach, Jorn
 APPLICANT: An, Jong-Qiang
 APPLICANT: Hamilton, Carol M.
 APPLICANT: Price, Jennifer L.
 APPLICANT: Raines, Tracy M.
 APPLICANT: Yu, Yang
 APPLICANT: Rameaka, Joshua G.
 APPLICANT: Page, Any
 APPLICANT: Matthew, Abraham V.
 APPLICANT: Ledford, Brooke L.
 APPLICANT: Woessner, Jeffrey P.
 APPLICANT: Haas, William David
 APPLICANT: Garcia, Carlos A.
 APPLICANT: Krieger, Maja
 APPLICANT: Slader, Ted
 APPLICANT: Davis, Keith R.
 APPLICANT: Allen, Keith
 APPLICANT: Hoffman, Neil
 APPLICANT: Hurban, Patrick
 TITLE OF INVENTION: Expressed Sequences of Arabidopsis
 FILE REFERENCE: 20230S (PARA-012PRV)
 CURRENT APPLICATION NUMBER: US/09/770,445
 CURRENT FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: US 60/178,472
 PRIOR FILING DATE: 2000-01-27

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:22:18 ; Search time 1366.46 Seconds
(without alignments)
15994.713 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750
Perfect score: 751
Sequence: 1 gaccaatgccagaatctc.....gactgaactgtcttgaaga 751

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.pl:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.ov:*
22: em.or:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pin:*
35: em.htg.rod:*
36: em.htg.man:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	751	100.0	10562	6	E38420	E38420 Novel polyp
2	751	100.0	170121	9	AF064860	AF064860 Homo sapi
3	751	100.0	340000	9	HS21C080	AL163280 Homo sapi
4	750.2	99.9	1565	9	AB041413	AB041413 Homo sapi
5	747.8	99.6	1576	9	AB041416	AB041416 Homo sapi
6	739.4	98.5	1360	9	AB041412	AB041412 Gorilla g
7	733.4	97.7	1570	9	AB041415	AB041415 Pan panis
8	721	96.0	1566	9	AB041414	AB041414 Pan trogl
9	710.8	94.6	1579	9	AB041417	AB041417 Pongo pyg
10	517.4	68.9	2762	9	AB020337	AB020337 Homo sapi
11	517.4	68.9	2775	6	E38419	E38419 Novel polyp
12	517	68.8	933	9	AF145784	AF145784 Homo sapi
13	508	67.6	933	9	HS6078	AT060678 Homo sapi
14	284.4	37.9	196900	2	AC020851	AC020851 Mus muscu
15	282.8	37.7	175861	2	AC120346	AF120346 Mus muscu
16	274.4	36.5	927	10	AF254738	AF254738 Mus muscu
17	132	17.6	917	10	AB039136	AB039136 Mus muscu
18	132	17.6	917	10	AB039137	AB039137 Mus muscu
19	132	17.6	917	10	AB039141	AB039141 Mus muscu
20	130.4	17.4	917	10	AB039138	AB039138 Mus muscu
21	130.4	17.4	917	10	AB039140	AB039140 Mus muscu
22	130.4	17.4	917	10	AB039142	AB039142 Mus muscu
23	130.4	17.4	917	10	AB039143	AB039143 Mus spici
24	128.8	17.2	917	10	AB039134	AB039134 Mus muscu
25	128.8	17.2	917	10	AB039135	AB039135 Mus muscu
26	128.8	17.2	917	10	AB039139	AB039139 Mus muscu
27	128.8	17.2	1175	10	AF029790	AF029790 Mus muscu
28	128.8	17.2	219368	10	AC098721	AC098721 Mus muscu
29	119.2	15.9	1037	9	AB041407	AB041407 Homo sapi
30	119.2	15.9	1053	9	AB041409	AB041409 Pan panis
31	119.2	15.9	1056	9	AB041410	AB041410 Gorilla g
32	119.2	15.9	1059	9	AB041411	AB041411 Pongo pyg
33	119.2	15.9	1060	9	AB041408	AB041408 Pan trogl
34	119.2	15.9	1739	6	E07739	E07739 cDNA encodi
35	119.2	15.9	1909	9	AF117222	AF117222 Homo sapi
36	119.2	15.9	202001	9	AC016723	AC016723 Homo sapi
37	91	12.1	996	9	AF494105	AF494105 Homo sapi
38	91	12.1	996	9	AF494106	AF494106 Homo sapi
39	91	12.1	996	9	HS6ALT4	Y15062 Homo sapien
40	91	12.1	1358	9	AB050856	AB050856 Homo sapi
41	91	12.1	1897	9	AB050855	AB050855 Homo sapi
42	91	12.1	2095	6	AX092304	AX092304 Sequence
43	91	12.1	2095	6	AX376094	AX376094 Sequence
44	91	12.1	2095	6	AX403321	AX403321 Sequence
45	91	12.1	2158	9	AK095917	AK095917 Homo sapi

ALIGNMENTS

RESULT 1	E38420	E38420	10562 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	E38420	Novel polypeptide.				
DEFINITION	E38420	Novel polypeptide.				
ACCESSION	E38420					
VERSION	E38420.1	GI:18626994				
KEYWORDS	JP 2000245464-A/2.					
SOURCE	Homo sapiens.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
TITLE	1 (bases 1 to 10562)					
JOURNAL	Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.					
	Novel polypeptide					
	Patent: JP 2000245464-A 2 12-SEP-2000;					

```
COMMENT
KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464 -A/2
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
C12P21/02,
PC C12P21/08, C12Q1/68, G01N33/53/(C12N1/21, C12R1/185), (C12N5/10,
C12R1/91),
PC (C12P21/02, C12R1/185), (C12P21/02, C12R1/91), C12N5/00, C12N5/00,
PC (C12N5/00, C12R1/91)
CC
FH Key Location/Qualifiers
FT promoter (1), (5000)
FT exon (5001), (5140)
FT exon (5001), (5273)
FT exon (5459), (5567)
FT exon (7427), (7586)
FT exon (8234), (10562)
FEATURES
Source 1..10562
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2610 a 2415 c 2574 g 2963 t
ORIGIN
Query Match 100.0% Score 751: DB 6: Length 10562;
Best Local Similarity 100.0% Pred. No. 4.5e-205; Mismatches 0; Indels 0; Gaps 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGACCAATGCCAGAAATCTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGC 60
Db 8000 GGACCAATGCCAGAAATCTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGC 8059
Qy 61 AGATGTTAGACCTTTGCTTACTGTTTAAACACACAGACCCGACTTCTGTATGCACG 120
Db 8060 AGATGTTAGACCTTTGCTTACTGTTTAAACACACAGACCCGACTTCTGTATGCACG 8119
Qy 121 GAGCTTCTAGAGTTTCCAAAACACGGGTCTCCTCTCCACCTCAGCCTCTAGCATAAAA 180
Db 8120 GAGCTTCTAGAGTTTCCAAAACACGGGTCTCCTCTCCACCTCAGCCTCTAGCATAAAA 8179
Qy 181 CTAGACATCCTCATGCTTTTGGGTCTAATCATGATTTGTTCTTTCAGATGGCT 240
Db 8180 CTAGACATCCTCATGCTTTTGGGTCTAATCATGATTTGTTCTTTCAGATGGCT 8239
Qy 241 TTCCCGAAGATGAGATTGATGATATTTGCTTCTGTTCTGGGGCTCTTTGTTGTAT 300
Db 8240 TTCCCGAAGATGAGATTGATGATATTTGCTTCTGTTCTGGGGCTCTTTGTTGTAT 8299
Qy 301 TTTAGCATGTACAGCTTAATCTTTCAAGACAGTCTTTGTTTACAGAAAGACGGG 360
Db 8300 TTTAGCATGTACAGCTTAATCTTTCAAGACAGTCTTTGTTTACAGAAAGACGGG 8359
Qy 361 AACTTCTTAAAGTCCAGATACAGACTGCAGGCAGACACCTCCCTTCTGCTGCTGCTG 420
Db 8360 AACTTCTTAAAGTCCAGATACAGACTGCAGGCAGACACCTCCCTTCTGCTGCTGCTG 8419
Qy 421 GTGACCTATCCCAACAAGTGTGGTGTAGCGGATGGCATCGGCAGACGCTGGGGGAAA 480
Db 8420 GTGACCTATCCCAACAAGTGTGGTGTAGCGGATGGCATCGGCAGACGCTGGGGGAAA 8479
Qy 481 GAGAGGATGTGAAGGAAAGAGCTGAAGACATTTCTCTCTGCGGACCCACGACGT 540
Db 8480 GAGAGGATGTGAAGGAAAGAGCTGAAGACATTTCTCTCTGCGGACCCACGACGT 8539
Qy 541 GCACGGAAACGAAGAGGTGGACCGAGGAGCCAGCAGCGGGGACATTTATCCAGAG 600
Db 8540 GCACGGAAACGAAGAGGTGGACCGAGGAGCCAGCAGCGGGGACATTTATCCAGAG 8599
```

```
Qy 601 GATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGATGGCATAGATGGGTC 660
Db 8600 GATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGATGGCATAGATGGGTC 8659
Qy 661 CATCGCTTTTGTCTCCTCAGCGGGGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 720
Db 8660 CATCGCTTTTGTCTCCTCAGCGGGGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 8719
Qy 721 GTTGACTATCTGACTGAATGCTTCTGCAAGA 751
Db 8720 GTTGACTATCTGACTGAATGCTTCTGCAAGA 8750
RESULT 2
AF064860 170121 bp DNA linear PRI 05-MAR-2002
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
sequence.
ACCESSION AF064860
VERSION AF064860.2 GI:18958624
KEYWORDS HTG: HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170121)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Schudy,A.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Hornischer,K., Brandt,P., Schaefer,M., Schoen,O., Desari,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Rieselmann,L., Dagand,E., Haaf,T., Wehrmeyer,H.,
Borzum,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
PUBMED 10830953
REFERENCE 2 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 170121)
REFERENCE Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 170121)
REFERENCE Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Feb 27, 2002 this sequence version replaced gi:3171153.
FEATURES
Location/Qualifiers
1..170121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 70124"
BASE COUNT 45184 a 36756 c 37509 g 50672 t
ORIGIN
```

```
Query Match 100.0% Score 751; DB 9; Length 170121;
Best Local Similarity 100.0%; Pred. No. 4.9e-205; Zimmels 0; Gaps 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0;

Qy 1 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTGCACATGCTGGC 60
Db 89814 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTGCACATGCTGGC 89873
Qy 61 AGATGTTAGACCTTTGTGCTTAACCTTTAAACACACACACACCGGCTTTCTGTATGCAGC 120
Db 89874 AGATGTTAGACCTTTGTGCTTTAACTTTAAACACACACACACCGGCTTTCTGTATGCAGC 89933
Qy 121 GAGTCTTACAGTTCCTCAACACAGGCTCTCTCTCCACCTCAGCCTCCTAGCATATAAA 180
Db 89934 GAGTCTTACAGTTCCTCAACACAGGCTCTCTCTCCACCTCAGCCTCCTAGCATATAAA 89993
Qy 181 CTAGACACATCCTCATGCTTTTGGAGTCTTAATCAATGTTGTTTCTCTTTCAGATGGCT 240
Db 89994 CTAGACACATCCTCATGCTTTTGGAGTCTTAATCAATGTTGTTTCTCTTTCAGATGGCT 90053
Qy 241 TTCCCGAAGATGAGATGATGATATTTGCTTCTGCTTCTGGGGCTTTTGTGTAT 300
Db 90054 TTCCCGAAGATGAGATGATGATATTTGCTTCTGCTTCTGGGGCTTTTGTGTAT 90113
Qy 301 TTTAGCATGTACAGTCTTAATCTTCAAGAACACAGTCTCTTTTACAAAGAACACGGG 360
Db 90114 TTTAGCATGTACAGTCTTAATCTTCAAGAACACAGTCTCTTTTACAAAGAACACGGG 90173
Qy 361 AACTTCCTTAAGTCCGACATGACAGTCTGAGGACACACCTCTCTCTCTCTCTCTCTG 420
Db 90174 AACTTCCTTAAGTCCGACATGACAGTCTGAGGACACACCTCTCTCTCTCTCTCTG 90233
Qy 421 GTGACCTTCATCCACAAACAGTTGGCTGAGCGCATGCCATCGGCGAGACGTGGGGAAA 480
Db 90234 GTGACCTTCATCCACAAACAGTTGGCTGAGCGCATGCCATCGGCGAGACGTGGGGAAA 90293
Qy 481 GAGAGGATGTTGAAGGAAAGCAGCTGAAGACATTTCTCTCTCTGGGACACACAGT 540
Db 90294 GAGAGGATGTTGAAGGAAAGCAGCTGAAGACATTTCTCTCTCTGGGACACACAGT 90353
Qy 541 GCAGCGAAACGAAAGAGTGGACAGGAGAGCCAGCAGCAGCGGACATTTATCCAGAG 600
Db 90354 GCAGCGAAACGAAAGAGTGGACAGGAGAGCCAGCAGCAGCGGACATTTATCCAGAG 90413
Qy 601 GATTTCCTAGACGCTCTATTACAATCTGACCTGACGACCATGATGGGCATAGAAATGGGTC 660
Db 90414 GATTTCCTAGACGCTCTATTACAATCTGACCTGACGACCATGATGGGCATAGAAATGGGTC 90473
Qy 661 CATCGCTTTTGTCTCAGCGGGCTTTGTGATGAACACAGCTCAGACATGTTTCATCAAT 720
Db 90474 CATCGCTTTTGTCTCAGCGGGCTTTGTGATGAACACAGCTCAGACATGTTTCATCAAT 90533
Qy 721 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 90534 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 90564

RESULT 3
HS21C080 340000 bp DNA linear PRI 24-MAY-2000
LOCUS Homo sapiens chromosome 21 segment HS21C080.
DEFINITION AL163280 AP001735 BA000005
ACCESSION AL163280.2 GI:7717369
VERSION 1 (bases 1 to 340000)
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
```

```
TITLE
JOURNAL

COMMENT
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gs.riken.go.jp
* URL: http://hgp.gs.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
Mascheröder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
Innistrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P16019, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF121897"
73410..243333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF064860"
224137..>340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF121782"
286628..>340000
/organism="Homo sapiens"
```

```
/db_xref="taxon.9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088..1566
/repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398..2495
/repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2496..2878
/repeat_region
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
2879..2931
/repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2932..3004
/repeat_region
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3005..3074
/repeat_region
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3075..3157
/gene="SH3BGR"
join(<3473..3565,12649..12678,15331..15460,18739..19157)
mRNA
/gene="SH3BGR"
join(3473..3565,12649..12678,15331..15426)
CDS
/partial
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/protein_id="CAB90445.1"
/db_xref="GI:7717370"
/translation="GSEKAEGGETEAQKESGSDVGNLPEAQKNEEGETAPETEE
IAMGAGCEAEETEAEGEPGEDEDS"
3473..3565
/exon
/gene="SH3BGR"
/number=4
5542..5829
/repeat_region
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6887..7067)
/repeat_region
/note="L1M3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8482..8762)
/repeat_region
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8865..8964)
/repeat_region
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
9131..9424
/repeat_region
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(9723..9882)
/repeat_region
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9989..10077)
/repeat_region
/note="L2"
```

```
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10078..10429)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
10980..11000
/repeat_region
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(11002..11282)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/gene="SH3BGR"
/number=5
complement(13115..13425)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
/repeat_region
/note="(CA)n"
/rpt_family="Simple_repeat"

Query Match 100.0% Score 751; DB 9; Length 340000;
Best Local Similarity 100.0% Pred. No. 5e-205;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTCAGATGCTGGC 60
|||||
DB 163222 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTCAGATGCTGGC 163281
|||||
QY 61 AGATGTTAGACCTTTGTGCTTAACCTGTTAAACACACAGACCCGACCTCTGTATGCAGC 120
|||||
DB 163282 AGATGTTAGACCTTTGTGCTTAACCTGTTAAACACACAGACCCGACCTCTGTATGCAGC 163341
|||||
QY 121 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCCCTCCTAGCATAAAA 180
|||||
DB 163342 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCCCTCCTAGCATAAAA 163401
|||||
QY 181 CTAGACACATCTCATGCTTTTGAGTCTTAATCATTTGATTTGCTTCTTTCAGATGGCT 240
|||||
DB 163402 CTAGACACATCTCATGCTTTTGAGTCTTAATCATTTGATTTGCTTCTTTCAGATGGCT 163461
|||||
QY 241 TTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGTTCTGGGGGCTCTTTGTTTGTAT 300
|||||
DB 163462 TTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGTTCTGGGGGCTCTTTGTTTGTAT 163521
|||||
QY 301 TTTAGCATGTACAGTCTAAATCCTTTTCAAGAACAGTCTCTTTGTTTCAAGAACAGACGGG 360
|||||
DB 163522 TTTAGCATGTACAGTCTAAATCCTTTTCAAGAACAGTCTCTTTGTTTCAAGAACAGACGGG 163581
|||||
QY 361 AACTTCTTAAAGCTCCCGATACAGACTGCAGGACAGACCTCCCTCTCTCTCTCTCTCTCTG 420
|||||
DB 163582 AACTTCTTAAAGCTCCCGATACAGACTGCAGGACAGACCTCCCTCTCTCTCTCTCTCTG 163641
|||||
```

Qy 421 GTACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAA 480
Db 163642 GTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAA 163701
Qy 481 GAGAGGATGTGAAGGAAGAGAGCTGAAGACATCTTCTCTCTGGGACACACAGCAGT 540
Db 163702 GAGAGGATGTGAAGGAAGAGAGCTGAAGACATCTTCTCTCTGGGACACACAGCAGT 163761
Qy 541 GCAGCGAAACGAAAGAGGTGGACAGGAGAGCGACGACGAGCGGGGACATTTCCAGAG 600
Db 163762 GCAGCGAAACGAAAGAGGTGGACAGGAGAGCGACGACGAGCGGGGACATTTCCAGAG 163821
Qy 601 GATTTCCTAGACGCTCTATTACATCTGACCCCTGAAGACCATGATGGCATGAATGGGTC 660
Db 163822 GATTTCCTAGACGCTCTATTACATCTGACCCCTGAAGACCATGATGGCATGAATGGGTC 163881
Qy 661 CATCGCTTTTGTCTCAGCGGGCGTTGTGATGAAACACAGACTCAGACATGTTCAAT 720
Db 163882 CATCGCTTTTGTCTCAGCGGGCGTTGTGATGAAACACAGACTCAGACATGTTCAAT 163941
Qy 721 GTTCACTATCTGACTGACTGCTCTCTGAAGA 751
Db 163942 GTTCACTATCTGACTGACTGCTCTCTGAAGA 163972

RESULT 4
AB041413
LOCUS Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc 1565 bp DNA linear PRI 13-APR-2000
DEFINITION beta1,3-galactosyltransferase 5, partial cds.
ACCESSION AB041413
VERSION AB041413.1 GI:7593020
KEYWORDS
SOURCE Homo sapiens (isolate:human-NR) DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1565)
Liu,Y. and Saitou,N.
Silver Project
Published Only in DataBase (2000)
2 (bases 1 to 1565)
Liu,Y. and Saitou,N.
Direct Submission
Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:msaitou@genes.nig.ac.jp,
URL:http://sayer.lab.nig.ac.jp/-silver/, Tel:81-559-81-6790,
Fax:81-559-81-6789)
Location/Qualifiers
1. .1565
/organism="Homo sapiens"
/isolate="human-NR"
/db_xref="taxon:9606"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..27
/number=3
28..674
/number=3
675..1565
/gene="beta1,3-GalT 5"
675..>1565
/gene="beta1,3-GalT 5"
/codon_start=1
/product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
/protein_id="BA94498.1"
/db_xref="GI:7593021"
/translation="NAFPMRLMYICLLVIGALCLYFSWYSLNPFKEOSFYVKXDNF
LKLPDQDCRQTPFVLVLTSSHKOLASERAIQTKERVYKQKLTFFELGTSS
AATKRVDESSEHGDIOKDFLDYNNLTLMGMIEWHRFCQAFVWKTDMDF
INVDYTELLKRNRTRETFGLKNEFFIPQPFKWFVSKSEYPPFPCSGTG
YVFSGVASQVYNVSKSVYIKLEDVFGLCLERLNIRLEELHSQTFPPGGLRFSVC

LFRRIVACHFIKPRLTWTGR*
675..>1565
/gene="beta1,3-GalT 5"
/number=4
BASE COUNT 410 a 370 c 385 g 398 t 2 others
ORIGIN
Query Match 99.9%; Score 750.2; DB 9; Length 1565;
Best Local Similarity 99.7%; Pred. No. 7.2e-205;
Matches 749; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGGC 60
Db 441 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGGC 500
Qy 61 AGATGTTAGACCTTTGTGCTTAACCTAAGTAAACACACAGACCCGACTTCTGTATGCAGC 120
Db 501 AGATGTTAGACCTTTGTGCTTAACCTAAGTAAACACACAGACCCGACTTCTGTATGCAGC 560
Qy 121 GAGGTTCTAGAGTTTCCAAACACAGGGGTCTCTCCACCTCAGCTCCTACGATAAAA 180
Db 561 GAGGTTCTAGAGTTTCCAAACACAGGGGTCTCTCCACCTCAGCTCCTACGATAAAA 620
Qy 181 CTAGACACATCCTCATGCTTTTTCAGGTCTAATCATTTGATTTTTCCTTTCAGATGGCT 240
Db 621 CTAGACACATCCTCATGCTTTTTCAGGTCTAATCATTTGATTTTTCCTTTCAGATGGCT 680
Qy 241 TTCCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGCTGGGGCTCTTTGTTGTAT 300
Db 681 TTCCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGCTGGGGCTCTTTGTTGTAT 740
Qy 301 TTTAGCATGTACAGTCTAAATCCCTTCAAGAACAGTCCTTTGTTTACAAAGAACAGCGG 360
Db 741 TTTAGCATGTACAGTCTAAATCCCTTCAAGAACAGTCCTTTGTTTACAAAGAACAGCGG 800
Qy 361 AACTTCCTTAAGTCTCCAGATACAGACTGAGGAGACACCTCCCTTCTCTGCTGCTGCTG 420
Db 801 AACTTCCTTAAGTCTCCAGATACAGACTGAGGAGACACCTCCCTTCTCTGCTGCTGCTG 860
Qy 421 GTGACCTCATCCCAACACAGTTGGCTGAGCGATGCCATCCGCGAGAGCTGGGGAAA 480
Db 861 GTGACCTCATCCCAACACAGTTGGCTGAGCGATGCCATCCGCGAGAGCTGGGGAAA 920
Qy 481 GAGAGGATGTGAAGGAAAGACAGCTGAAGACATTTCTCTCTCTGGGGACACACAGCAGT 540
Db 921 GAGAGGATGTGAAGGAAAGACAGCTGAAGACATTTCTCTCTCTGGGGACACACAGCAGT 980
Qy 541 GCAGCGAAACGAAAGAGGTGGACAGGAGCGACGACGAGCGGGGACATTTATCCAGAG 600
Db 981 GCAGCGAAACGAAAGAGGTGGACAGGAGCGACGAGCGGGGACATTTATCCAGAG 1040
Qy 601 GATTTCCTAGACGCTCTATTACATCTGACCCCTGAAGACCATGATGGCATGAATGGGTC 660
Db 1041 GATTTCCTAGACGCTCTATTACATCTGACCCCTGAAGACCATGATGGCATGAATGGGTC 1100
Qy 661 CATCGCTTTTGTCTCAGCGGGCGTTTGTGATGAAACACACACTCAGACATGTTCAAT 720
Db 1101 CATCGCTTTTGTCTCAGCGGGCGTTTGTGATGAAACACACACTCAGACATGTTCAAT 1160
Qy 721 GTTCACTATCTGACTGACTGCTCTCTGAAGA 751
Db 1161 GTTCACTATCTGACTGACTGCTCTCTGAAGA 1191

RESULT 5
AB041416
LOCUS Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc 1576 bp DNA linear PRI 17-OCT-2000
DEFINITION beta1,3-galactosyltransferase 5, partial cds.
ACCESSION AB041416
VERSION AB041416.1 GI:7593026
KEYWORDS
SOURCE Homo sapiens (isolate:#056) DNA.

```

ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Liu,Y. and Saitou,N.
TITLE        Silver Project
JOURNAL       Published Only in DataBase (2000)
REFERENCE
AUTHORS      Liu,Y. and Saitou,N.
TITLE        Direct Submission
JOURNAL       Submitted (11-APR-2000) Naruya Saitou, National Institute of
               Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
               Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
               URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
               Fax:81-559-81-6789)
FEATURES
   source     Location/Qualifiers
               1..1576
               /organism="Homo sapiens"
               /isolate="#056"
               /db_xref="taxon:9606"
               /note="human sequence used for primer design based on Acc#
               AB020337"
               <1..27
               /number=3
               28..674
               /number=3
               675..1576
               /gene="betal,3-Galt 5"
               675..>1576
               /gene="betal,3-Galt 5"
               /codon_start=1
               /product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
               /protein_id="BAA94501.1"
               /db_xref="GI:7593027"
               /translation="MAFPMRLMVICLLVLGALCLYFSMYSLNPKSQSFYKKDGNF
               LKPTDCRQPPFVLVLTSSHKQALRMALROTGWKERTVKGKQLKTFLLGTTSS
               RAKTEVQESQRHGDIIQKDFLDVYVNLTKTMGLEWHRFQCPQAFVWKDSDMF
               INVYDLTELLKKNTRTFRTFGFLKNEFPFQFVSKSFVSKSYWDRPVPCCSGG
               YVFGDVASQVFNVSQVYFKLEDFVGLCLERLNRLLELHQSPTFFFGGURFSVC
               LPRFVACHFTKPRLLDYQWALE"
               675..>1576
               /gene="betal,3-Galt 5"
               /number=4
               373 C 387 G 401 T
BASE COUNT   415 A 373 C 387 G 401 T
ORIGIN
Query Match   99.6%; Score 747.8; DB 9; Length 1576;
Best Local Similarity 99.7%; Pred. No. 3.5e-204;
Matches 749; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTGACAGATCTGCG 60
    |||
DB 441 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTGACAGATCTGCG 500
    |||
QY 61 AGATGTTAGACCTTTGTGCTTAACCTTTTAAACACACAGACCCGACTTCTGTATGCAGC 120
    |||
DB 501 AGATGTTAGACCTTTGTGCTTAACCTTTTAAACACACAGACCCGACTTCTGTATGCAGC 560
    |||
QY 121 GAGTTCCTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCCTTAGCATATAA 180
    |||
DB 561 GAGTTCCTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCCTTAGCATATAA 620
    |||
QY 181 CTAGACATCTCTATGCTTTTGTAGCTTAATCATGTGATTTTGTCTTTTCAGATGGCT 240
    |||
DB 621 CTAGACATCTCTATGCTTTTGTAGCTTAATCATGTGATTTTGTCTTTTCAGATGGCT 680
    |||
QY 241 TTCCCGAAGATGAGATGATGATATTCGCTTCTGTTCTGGGCTCTTTGTTGTAT 300
    |||
DB 681 TTCCCGAAGATGAGATGATGATATTCGCTTCTGTTCTGGGCTCTTTGTTGTAT 740
    |||
QY 301 TTGTAGCTGTACAGTCTAAATCTTTCAAGAACAGCTCTTTGTTTACAAGAAAGACGGG 360
    |||
DB 741 TTGTAGCTGTACAGTCTAAATCTTTCAAGAACAGCTCTTTGTTTACAAGAAAGACGGG 800

```

```

QY 361 AACTTCCTTAAGCTCCAGATACAGACTCAGGCAGACACCTCCCTCTCTCTCTGCTGCTG 420
    |||
DB 801 AACTTCCTTAAGCTCCAGATACAGACTCAGGCAGACACCTCCCTCTCTCTCTGCTGCTG 860
    |||
QY 421 GTGACCTCATCCCAACAAAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGGGAAA 480
    |||
DB 861 GTGACCTCATCCCAACAAAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGGGAAA 920
    |||
QY 481 GAGAGGATGGTGAAGGGAAGAGCTGAAGACATTTCTCTCTGGGGACACGACGAGT 540
    |||
DB 921 GAGAGGATGGTGAAGGGAAGAGCTGAAGACATTTCTCTCTGGGGACACGACGAGT 980
    |||
QY 541 GCAGCGGAAACGAAAGAGGTGGACACGAGAGAGCCAGCGGACGAGGAGGACATTTATCCAGAAG 600
    |||
DB 981 GCAGCGGAAACGAAAGAGGTGGACACGAGAGAGCCAGCGGACGAGGACATTTATCCAGAAG 1040
    |||
QY 601 GATTTCCTAGACGCTATTACAACTGTGACCTGGAAGACCATGATGGGCATAGAAATGGGTC 660
    |||
DB 1041 GATTTCCTAGACGCTATTACAACTGTGACCTGGAAGACCATGATGGGCATAGAAATGGGTC 1100
    |||
QY 661 CATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAAT 720
    |||
DB 1101 CATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAAT 1160
    |||
QY 721 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
    |||
DB 1161 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 1191
    |||
RESULT 6
AB041412 1360 bp DNA linear PRI 17-OCT-2000
LOCUS      Gorilla gorilla betal,3-Galt 5 gene for UDP-Gal:GlcNAc
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041412
VERSION    AB041412.1 GI:7593018
KEYWORDS   Gorilla gorilla (isolate:#085) DNA.
SOURCE     Gorilla gorilla
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE  1 (bases 1 to 1360)
AUTHORS    Liu,Y. and Saitou,N.
TITLE      Silver Project
JOURNAL     Published Only in DataBase (2000)
REFERENCE  2 (bases 1 to 1360)
AUTHORS    Liu,Y. and Saitou,N.
TITLE      Direct Submission
JOURNAL     Submitted (11-APR-2000) Naruya Saitou, National Institute of
               Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
               Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
               URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
               Fax:81-559-81-6789;
               Location/Qualifiers
               1..1360
               /organism="Gorilla gorilla"
               /isolate="#085"
               /db_xref="taxon:9593"
               /note="human sequence used for primer design based on Acc#
               AB020337"
               <1..466
               /number=3
               467..1360
               /gene="betal,3-Galt 5"
               487..>1360
               /gene="betal,3-Galt 5"
               /codon_start=1
               /product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
               /protein_id="BAA94497.1"
               /db_xref="GI:7593019"
               /translation="MAFPMRLMVICLLVLGALCLYFSMYSLNPKSQSFYKKDGNF
               LKLPDTCRQPPFVLVLTSSHKQALRMALROTGWKERTVKGKQLKTFLLGTTSS

```

AAETKEVDSESRHGDIIQKDFLDVYVNLTKTMGIEWHVFPCPOAFVWKTDSDMF
INVDYLTLLKNNRTTFFGFLKLNFFPIQPFKSFVSKSEYVWDRYPPFCGSGTG
YFSGDVASQVNVSESVYIKLEDYFVGLCLERLNIRLELHSHQPTFFPGLRFSVC
RFRRIIVACHFIKPRLLDYWQA*
467. >1360
/gene="beta1.3-Galt 5"
/number=4

BASE COUNT 334 a 331 c 341 g 351 t 3 others
ORIGIN

Query Match 98.5%; Score 739.4; DB 9; Length 1360;
Best Local Similarity 98.9%; Pred. No. 9.2e-202;
Matches 743; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GGACCAATGCCAGACTCTGAGCTTTTATCTTACACATCAAGTGACAGATGCTGGC 60
Db 233 GGCAATAGCCAGAACTCTGAGCTTTTATCTTACACATCAAGTGACAGATGCTGGC 292
Qy 61 AGATGTTAGACCTTTGCTTAACTGTTTAAACCACACAGACCCGACTTCTGTATGCAGC 120
Db 293 AGATGTTAGACCTTTGCTTAACTGTTTAAACCACACAGACCCGACTTCTGTATGCAGC 352
Qy 121 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCTCTCTAGCATATAAA 180
Db 353 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCTCTCTAGCATATAAA 412
Qy 181 CTAGACACATCCTCATCTGCTTTGAGGTCTAATCATTTGATTTGTTTCTCTTTCAGATGGCT 240
Db 413 CTAGACACATCCTCATCTGCTTTGAGGTCTAATCATTTGATTTGTTTCTCTTTCAGATGGCT 472
Qy 241 TTCCCGAAGATGAGATGATGATATATTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 473 TTCCCGAAGATGAGATGATGATATATTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 532
Qy 301 TTTAGCTGTACAGTCTAAATCTTCAAGAACAGTCTCTTCTTTACAGAAAGACGGG 360
Db 533 TTTAGCTGTACAGTCTAATCTTCAAGAACAGTCTCTTCTTTACAGAAAGACGGG 592
Qy 361 AACTTCCTTAAGCTCCAGATACAGATGCTGAGGACAGACCTCCCTTCCTGCTGCTGCTGCTG 420
Db 593 AACTTCCTTAAGCTCCAGATACAGATGCTGAGGACAGACCTCCCTTCCTGCTGCTGCTGCTG 652
Qy 421 GTGACCTCATCCCAACAGTGTGGCTGAGCGATGCCATCCGACGAGCTGGGGGAAA 480
Db 653 GTGACCTCATCCCAACAGTGTGGCTGAGCGATGCCATCCGACGAGCTGGGGGAAA 712
Qy 481 GAGAGGATGTGAAGGAAAGACAGCTGAAGACATTTCTTCTGCTGGGACCCACAGCAGT 540
Db 713 GAGAGGACAGTGAAGGAAAGACAGCTGAAGACATTTCTTCTGCTGGGACCCACAGCAGT 772
Qy 541 GCACGGAAAGAGAGTGGACAGGAGAGCCAGGACGAGCCAGGACGAGCTATTCAGAGAG 600
Db 773 GCACGGAAAGAGAGTGGACAGGAGAGCCAGGAGCCAGGACGAGCTATTCAGAGAG 832
Qy 601 GATTTCCTAGACGCTCTATTACAATCTGACCTGAAGACCATGATGGCATAGAGTGGGTC 660
Db 833 GATTTCCTAGAGTCTATTACAATCTGACCTGAAGACCATGATGGCATAGAGTGGGTC 892
Qy 661 CATCGCTTTGCTCTAGGCGGGTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 720
Db 893 CATCGCTTTGCTCTAGGCGGGTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 952
Qy 721 GTTGACTATCTGACTGAAGTCTTCTGAAGA 751
Db 953 GTTGACTATCTGACTGAAGTCTTCTGAAGA 983

RESULT 7
AB041415 1570 bp DNA linear PRI 13-APR-2000
LOCUS Pan paniscus beta1.3-Galt 5 gene for UDP-Gal:GlcNAc
DEFINITION beta1.3-galactosyltransferase 5, partial cds.
ACCESSION AB041415

VERSION AB041415.1 GI:7593024
KEYWORDS Pan paniscus (isolate:bonobo-05) DNA.
SOURCE Pan paniscus
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1570)
AUTHORS Liu, Y. and Saitou, N.
TITLE Silver Project
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1570)
AUTHORS Liu, Y. and Saitou, N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genetics.nig.ac.jp,
URL: http://sayer.lab.nig.ac.jp/~silver/, Tel: 81-559-81-6790,
Fax: 81-559-81-6789)
FEATURES
source
1..1570
Location/Qualifiers
/organism="Pan paniscus"
/isolate="bonobo-05"
/db_xref="taxon:9597"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..20
/number=3
21..667
/number=3
668..1570
/gene="beta1.3-Galt 5"
668..>1570
/gene="beta1.3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
/protein_id="BAA94500.1"
/db_xref="GI:7593025"
/translation="MAPKMLMYVCLLVGLALCYFSMYSLNLFKEQSFVYKDGNF
LKLPDRCQTPPLVLLVLTSSHKQLAERMAIQTWKGRTVKGKLTFTLLGTSS
AAETKEVDSESRHGDIIQKDFLDVYVNLTKTMGIEWHVFPCPOAFVWKTDSDMF
INVDYLTLLKNNRTTFFGFLKLNFFPIQPFKSFVSKSEYVWDRYPPFCGSGTG
YFSGDVASQVNVSESVYIKLEDYFVGLCLERLNIRLELHSHQPTFFPGLRFSVC
RFRRIIVACHFIKPRLLDYWALEN"
668..>1570
/gene="beta1.3-Galt 5"
/number=4
408 a 372 c 392 g 398 t

BASE COUNT 408 a 372 c 392 g 398 t
ORIGIN

Query Match 97.7%; Score 733.4; DB 9; Length 1570;
Best Local Similarity 98.5%; Pred. No. 5e-200;
Matches 740; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACATCAAGTGACAGATGCTGGC 60
Db 434 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACATCAAGTGACAGATGCTGGC 493
Qy 61 AGATGTTAGACCTTTGCTTAACTGTTTAAACCACACAGACCCGACTTCTGTATGCAGC 120
Db 494 AGATGTTAGACCTCTGCTTAACTGTTTAAACCACACAGACCCGACTTCTGTATGCAGC 553
Qy 121 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCTCTCTAGCATATAAA 180
Db 554 GAGGTTCTAGAGTTTCCAAAACACGGGTCTCTCTCCACCTCAGCTCTCTAGCATATAAA 613
Qy 181 CTAGACACATCCTCATCTGCTTTGAGGTCTAATCATTTGCTTCTCTTTCAGATGGCT 240
Db 614 CTAGACACATCCTCATCTGCTTTGAGGTCTAATCATTTGCTTCTCTTTCAGATGGCT 673
Qy 241 TTCCCGAAGATGAGATGATGATATATTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 674 TTCCCGAAGATGAGATGATGATATATTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733

AB041417 1579 bp DNA linear PRI 13-APR-2000
LOCUS Pongo pygmaeus betal,3-Galt 5 gene for UDP-Gal:GlcNac
DEFINITION betal,3-galactosyltransferase 5, partial cds.
ACCESSION AB041417.1 GI:7593028
VERSION
KEYWORDS
SOURCE
ORGANISM
Pongo pygmaeus (isolate:oran-pol7) DNA.
Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pongo.
REFERENCE
AUTHORS Liu, Y. and Saitou, N.
TITLE Silver Project
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 1579)
AUTHORS Liu, Y. and Saitou, N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL:htp://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
Fax:81-559-81-6789)
FEATURES
Source
Location/Qualifiers
1..1579
/organism="Pongo pygmaeus"
/isolate="oran-Pol7"
/db_xref="taxon:9600"
/note="human sequence used for primer design based on Acc#
AB020337"
exon <1..26
intron 27..673
gene 674..1579
CDS 674..>1579
/gene="betal,3-Galt 5"
/gene="betal,3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNac betal,3-galactosyltransferase 5"
/protein_id="BAA94502.1"
/db_xref="GI:7593029"
/translation="MAFPMRLIYICLLVGLCALCYFSMNLNPFQSVYKKEGN
FLKPDTCRTPPLVLLVSSHQLAERMAIRQTGWKERTVKGOLKRTFFLLGTS
SAAEKQDSQSRHGDIIQKDLVDVYNLTLMKMGWVHRFCPOAARFVMTDSM
FINVDYELLKKNRTRFFGLKLNELPIROPFSKFWKSEYPMWRYPFPCST
AYVSGSDVSOVYNVSVRYKLEDFVGLCLERLIRLEELHSQPTFFPEGLRSV
CFRRIRVACHEIKPRTLLDYWQLEN"
exon 674..>1579
/gene="betal,3-Galt 5"
/number=4
BASE COUNT 421 a 371 c 383 g 404 t
ORIGIN
Query Match 94.6%; Score 710.8; DB 9; Length 1579;
Best Local Similarity 97.3%; Pred. No. 1.6e-193;
Matches 734; Conservative 0; Mismatches 17; Indels 3; Gaps 1;
QY 1 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGGC 60
DB 440 GGACCAATGCCAGATCTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGGC 499
QY 61 AGATGTTAGACCTTTGCTTAACTGTTTAAACACACAGACCCGACCTCTGTATGCAGC 120
DB 500 AGATGTTAGACCTCTGTGTTAACTGTTTAAACACACAGACCCGACCTCTGTATGCAGC 559
QY 121 GAGGTTCTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCCTCCTAGCATAAA 180
DB 560 GAGATTTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCCTCCTAGCATAAA 619
QY 181 CTAGACACATCCCATGCTTTTGTGATCTAAATCATGTTGATTTGCTTTCAGATGGCT 240
DB 620 CTAGACACATCCCATGCTTTTGTGATCTAAATCATGTTGATTTGCTTTCAGATGGCT 679

QY 241 TTCCCCAAGATCAGATTGATATATTTCCTCTGTTCTGGGGCTCTTTGTTGTTAT 300
DB 680 TTCCCCAAGATCAGATTGATATATTTCCTCTGTTCTGGGGCTCTTTGTTGTTAT 739
QY 301 TTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCTCTTTGTTTACAAGAAA---GAC 357
DB 740 TTAGCATGTACAACTGAATCCTTTCAAAGAACAGTCTCTTTGTTTACAAGAACAGGAC 799
QY 358 GGAATCTCCTTAAGTCCCATACAGATCAGAGTCCAGGACACACCTCCCTTCTGCTCTG 417
DB 800 GGAATCTCCTTAAGTCCCATACAGATCAGAGTCCAGGACACACCTCCCTTCTGCTCTG 859
QY 418 CTGCTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGACACAGTGGGG 477
DB 860 CTGCTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGACACAGTGGGG 919
QY 478 AAAGAGGAGTGGTGAAGGAAAGCAGCTGAAGACATTTCTTCTCTCTGGGACACACG 537
DB 920 AAAGAGGAGTGGTGAAGGAAAGCAGCTGAAGACATTTCTTCTCTCTGGGACACACG 979
QY 538 AGTGACAGGAAACAGAAAGAGTGGACAGGAGAGCCAGGACACGGGACATTTATCCAG 597
DB 980 AGTGACAGGAAACAGAAAGAGTGGACAGGAGAGCCAGGACACGGGACATTTATCCAG 1039
QY 598 AAGGATTTCTAGACGTCTATTACATCTGACCTGAAGACCATGATGGCATAGATGG 657
DB 1040 AAGGATTTCTAGACGTCTATTACATCTGACCTGAAGACCATGATGGCATAGATGG 1099
QY 658 GTCCATCGCTTTTGTCTCAGCGGGCTTTGTGATGAAGAACAGATCAGACATGTTTCATC 717
DB 1100 GTCCATCGCTTTTGTCTCAGCGGGCTTTGTGATGAAGAACAGATCAGACATGTTTCATC 1159
QY 718 AATGTTGACTATCTGACTGAATGCTGTTCTGAAGA 751
DB 1160 AATGTTGACTATCTGACTGAATGCTGTTCTGAAGA 1193
RESULT 10
AB020337
LOCUS Homo sapiens mRNA for UDP-Gal:GlcNac betal,3-galactosyltransferase
DEFINITION 5, complete cds.
ACCESSION AB020337
VERSION UDP-Gal:GlcNac betal,3-galactosyltransferase 5.
KEYWORDS UDP-Gal:GlcNac betal,3-galactosyltransferase 5.
SOURCE Homo sapiens Adenocarcinoma cell_line:Colo 205 cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isshiki, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M.,
Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T. and
Narimatsu, H.
TITLE Cloning, expression, and characterization of a novel
UDP-galactose:beta-N-acetylglucosamine
betal,3-galactosyltransferase (beta3Gal-T5) responsible for
synthesis of type 1 chain in colorectal and pancreatic epithelia
and tumor cells derived therefrom
J. Biol. Chem. 274 (18), 12499-12507 (1999)
99230269
2 (bases 1 to 2762)
Isshiki, S., Togayachi, A. and Narimatsu, H.
Direct Submission
TITLE Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University,
Institute of Life Science; 1-236, Tangi-cho, Hachioji, Tokyo
192-8577, Japan (E-mail:sisshiki@epo.lijnet.or.jp,
Tel:81-426-91-9456, Fax:81-426-91-9315)
LOCATION/Qualifiers
1..2762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"

FEATURES	source	Location/Qualifiers	1. .2775	2. .2775	3. .2775	4. .2775	5. .2775	6. .2775	7. .2775	8. .2775	9. .2775	10. .2775	11. .2775	12. .2775	13. .2775	14. .2775	15. .2775	16. .2775	17. .2775	18. .2775	19. .2775	20. .2775	21. .2775	22. .2775	23. .2775	24. .2775	25. .2775	26. .2775	27. .2775	28. .2775	29. .2775	30. .2775	31. .2775	32. .2775	33. .2775	34. .2775	35. .2775	36. .2775	37. .2775	38. .2775	39. .2775	40. .2775	41. .2775	42. .2775	43. .2775	44. .2775	45. .2775	46. .2775	47. .2775	48. .2775	49. .2775	50. .2775	51. .2775	52. .2775	53. .2775	54. .2775	55. .2775	56. .2775	57. .2775	58. .2775	59. .2775	60. .2775	61. .2775	62. .2775	63. .2775	64. .2775	65. .2775	66. .2775	67. .2775	68. .2775	69. .2775	70. .2775	71. .2775	72. .2775	73. .2775	74. .2775	75. .2775	76. .2775	77. .2775	78. .2775	79. .2775	80. .2775	81. .2775	82. .2775	83. .2775	84. .2775	85. .2775	86. .2775	87. .2775	88. .2775	89. .2775	90. .2775	91. .2775	92. .2775	93. .2775	94. .2775	95. .2775	96. .2775	97. .2775	98. .2775	99. .2775	100. .2775	101. .2775	102. .2775	103. .2775	104. .2775	105. .2775	106. .2775	107. .2775	108. .2775	109. .2775	110. .2775	111. .2775	112. .2775	113. .2775	114. .2775	115. .2775	116. .2775	117. .2775	118. .2775	119. .2775	120. .2775	121. .2775	122. .2775	123. .2775	124. .2775	125. .2775	126. .2775	127. .2775	128. .2775	129. .2775	130. .2775	131. .2775	132. .2775	133. .2775	134. .2775	135. .2775	136. .2775	137. .2775	138. .2775	139. .2775	140. .2775	141. .2775	142. .2775	143. .2775	144. .2775	145. .2775	146. .2775	147. .2775	148. .2775	149. .2775	150. .2775	151. .2775	152. .2775	153. .2775	154. .2775	155. .2775	156. .2775	157. .2775	158. .2775	159. .2775	160. .2775	161. .2775	162. .2775	163. .2775	164. .2775	165. .2775	166. .2775	167. .2775	168. .2775	169. .2775	170. .2775	171. .2775	172. .2775	173. .2775	174. .2775	175. .2775	176. .2775	177. .2775	178. .2775	179. .2775	180. .2775	181. .2775	182. .2775	183. .2775	184. .2775	185. .2775	186. .2775	187. .2775	188. .2775	189. .2775	190. .2775	191. .2775	192. .2775	193. .2775	194. .2775	195. .2775	196. .2775	197. .2775	198. .2775	199. .2775	200. .2775	201. .2775	202. .2775	203. .2775	204. .2775	205. .2775	206. .2775	207. .2775	208. .2775	209. .2775	210. .2775	211. .2775	212. .2775	213. .2775	214. .2775	215. .2775	216. .2775	217. .2775	218. .2775	219. .2775	220. .2775	221. .2775	222. .2775	223. .2775	224. .2775	225. .2775	226. .2775	227. .2775	228. .2775	229. .2775	230. .2775	231. .2775	232. .2775	233. .2775	234. .2775	235. .2775	236. .2775	237. .2775	238. .2775	239. .2775	240. .2775	241. .2775	242. .2775	243. .2775	244. .2775	245. .2775	246. .2775	247. .2775	248. .2775	249. .2775	250. .2775	251. .2775	252. .2775	253. .2775	254. .2775	255. .2775	256. .2775	257. .2775	258. .2775	259. .2775	260. .2775	261. .2775	262. .2775	263. .2775	264. .2775	265. .2775	266. .2775	267. .2775	268. .2775	269. .2775	270. .2775	271. .2775	272. .2775	273. .2775	274. .2775	275. .2775	276. .2775	277. .2775	278. .2775	279. .2775	280. .2775	281. .2775	282. .2775	283. .2775	284. .2775	285. .2775	286. .2775	287. .2775	288. .2775	289. .2775	290. .2775	291. .2775	292. .2775	293. .2775	294. .2775	295. .2775	296. .2775	297. .2775	298. .2775</
----------	--------	---------------------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	--------------

```

Db 820 AATGGGTCATCGCTTTTGTCTCAGCGGGGTTTGTGATGAAACAGACTCAGACATGT 879
Qy 713 TCATCAATGTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 880 TCATCAATGTGACTATCTGACTGAACCTGCTTCTGAAGA 918

RESULT 12
AF145784 AF145784 933 bp DNA linear PRI 14-NOV-1999
LOCUS Homo sapiens beta1,3 galactosyltransferase-V (B3GALT5) gene,
DEFINITION complete cds.
ACCESSION AF145784
VERSION AF145784.1 GI:6409192
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 933)
AUTHORS Zhou,D., Berger,E.G. and Hennet,T.
TITLE Molecular cloning of a human UDP-galactose:GlcNAc6beta1,3galNAC
beta1,3 galactosyltransferase gene encoding an O-linked
core3-elongation enzyme
JOURNAL Eur. J. Biochem. 263 (2), 571-576 (1999)
MEDLINE 99337698
PUBMED 10406968
REFERENCE 2 (bases 1 to 933)
AUTHORS Zhou,D. and Hennet,T.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1999) Physiology, University of Zurich,
Winterthurerstrasse 190, Zurich 8057, Switzerland
FEATURES
source
1..933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chrnosome="21"
/map="21q22"
1..933
/gene="B3GALT5"
1..933
/gene="B3GALT5"
/notes="glycosyltransferase"
/codon_start=1
/product="beta1,3 galactosyltransferase-V"
/protein_id="AAF07880.1"
/db_xref="GI:6409193"
/translation="MAFPKRLMYICLLVLGALCLYFSMISLPNPKESFYVKKDNF
LKLPDRCQTPPLVLLVTSSSHQKLAERMAIQTWKERTVKGKQLKTFLLGTTSS
AAETKEVDQESORHGDIQKDFLDVYINLTAKTMGIEWHRECPQAAFYVKMTSDMF
INVDYLTLLKKNRTRFTGFLKNEFPPIQPFKSWFVSKSEYPDWRYPPFCSTG
YVFGSDVASQVYNVSKSPYIKLEDYFVGLCLERLNIRLELHSQTFPFPGGLRFSVC
LFRRIVACHFIKPRILLDIWQALENSRGDCPPV"
BASE COUNT 227 a 232 c 241 g 233 t
ORIGIN
Query Match 68.8%; Score 517; DB 9; Length 933;
Best Local Similarity 100.0%; Pred. No. 1,1e-137;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 ATGGCTTCCCGAAGATGAGATGATATATTGCTTGGTCTGGGGGCTCTTTGT 294
Db 1 ATGGCTTCCCGAAGATGAGATGATATATTGCTTGGTCTGGGGGCTCTTTGT 60

Qy 295 TTGTATTATTAGCATGTACAGTCTAAATCCCTTCAAGAACAGTCCTTTGTATACAGAAA 354
Db 61 TTGTATTATTAGCATGTACAGTCTAAATCCCTTCAAGAACAGTCCTTTGTATACAGAAA 120

Qy 355 GACGGGAACCTTCCTTAAGCTCCAGATACAGACTGACGAGGACACCTCCCTTCCTCGTC 414
Db 121 GACGGGAACCTTCCTTAAGCTCCAGATACAGACTGACGAGGACACCTCCCTTCCTCGTC 180

Qy 415 CTGCTGGTGACCTCATCCCAACACAGTTGGCTCAGCGCATGCCATCCGGGACAGCTGG 474

```

```

Db 181 CTGCTGTGACCTCATCCCAACAAAGTTGGCTGAGCGATGCCATCCGGCAGACGTGG 240
Qy 475 GGGAAAGAGAGGATGGTGAAGGGAACAGCAGCTGAAGACATCTTCTCTCGGGGACACC 534
Db 241 GGGAAAGAGAGGATGGTGAAGGGAACAGCAGCTGAAGACATCTTCTCTCGGGGACACC 300
Qy 535 AGCAGTGCAGCGGAACGAAAGAGAGTGGACCGAGAGAGCCAGGACACGCGGACATTATC 594
Db 301 AGCAGTGCAGCGGAACGAAAGAGAGTGGACCGAGAGAGCCAGGACACGCGGACATTATC 360
Qy 595 CAGAAGGATTTCTPAGACGTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 654
Db 361 CAGAAGGATTTCTPAGACGTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 420
Qy 655 TGGTCCATCGCTTTTCTCTCAGCGGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 714
Db 421 TGGTCCATCGCTTTTCTCTCAGCGGCGCTTTGTGATGAAACAGACTCAGACATGTTTC 480
Qy 715 ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 481 ATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGA 517

RESULT 13
HSA6078 HSA6078 933 bp DNA linear PRI 11-MAY-2000
LOCUS Homo sapiens beta3gal-T6 gene.
DEFINITION
ACCESSION AJ006078
VERSION AJ006078.1 GI:7799922
KEYWORDS Beta-1,3-galactosyltransferase; beta3gal-T6 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 933)
AUTHORS Amado,M., Carneiro,F. and Clausen,H.
TITLE Cloning and expression of two beta-1,3-galactosyltransferases:
beta3gal-T5 and beta3gal-T6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 933)
AUTHORS Amado,M.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998) Amado M., Department of Oral Diagnostics,
Royal dental School, Norre Alle 20, 2200 Copenhagen, DENMARK
FEATURES
source
1..933
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..933
/gene="beta3gal-T6"
1..933
/gene="beta3gal-T6"
/codon_start=1
/product="beta-1,3-galactosyltransferase"
/protein_id="CAB91547.1"
/db_xref="GI:7799923"
/translation="MAFPKRLMYICLLVLGALCLYMSNLPNPKESFYVKKDNF
LKLPDRCQTPPLVLLVTSSSHQKLAERMAIQTWKERTVKGKQLKTFLLGTTSS
AAETKEVDQESORHGDIQKDFLDVYINLTAKTMGIEWHRECPQAAFYVKMTSDMF
INVDYLTLLKKNRTRFTGFLKNEFPPIQPFKSWFVSKSEYPDWRYPPFCSTG
YVFGSDVASQVYNVSKSPYIKLEDYFVGLCLERLNIRLELHSQTFPFPGGLRFSVC
LFRRIVACHFIKPRILLDIWQALENSRGDCPPV"
BASE COUNT 229 a 234 c 240 g 229 t
ORIGIN
Query Match 67.6%; Score 508; DB 9; Length 933;
Best Local Similarity 98.8%; Pred. No. 4.4e-135;
Matches 511; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 235 ATGGCTTCCCGAAGATGAGATGATATATTGCTTGGTCTGGGGGCTCTTTGT 294
Db 1 ATGGCTTCCCGAAGATGAGATGATATATTGCTTGGTCTGGGGGCTCTTTGT 60

```

```
QY 295 TTGATTTTAGCATGTACAGTCTAATCCTTTCRAAGACAGTCTTGTGTTTACAAGAA 354
|||||
Db 61 TTGATTTAGCATGAACAGTCTAATCCTTTCRAAGACAGTCTTGTGTTTACAAGAA 120
|||||
QY 355 GAGGGAACCTCTTAAAGTCCAGATACAGACTGCAGGCGAGACACCTTCTCTGTC 414
|||||
Db 121 GAGGGAACCTCTTAAAGTCCAGATACAGACTGCAGGCGAGACACCTTCTCTGTC 180
|||||
QY 415 CTGCTGGTACCTCATCCACAAACAGTTGGCTGAGCCATGGCCATCGGCACAGCTGG 474
|||||
Db 181 CTGCTGGTACCTCATCCACAAACAGTTGGCTGAGCCATGGCCATCGGCACAGCTGG 240
|||||
QY 475 GGAAGAGAGAGTGTGAAGGAAAGAGCAGCTGAAGACATTTCTTCTCTGGGACACAC 534
|||||
Db 241 GGAAGAGAGAGTGTGAAGGAAAGAGCAGCTGAAGACATTTCTTCTCTGGGACACAC 300
|||||
QY 535 AGCAGTGCAGGGAAGAGAGAGTGGACAGGAGAGCCAGGACGACGCGGAGCATATTC 594
|||||
Db 301 AGCAGTGCAGGGAAGAGAGAGTGGACAGGAGAGCCAGGACGACGCGGAGCATATTC 360
|||||
QY 595 CAGAAGGATTTCTAGACGCTTATTACAACTCTGACCTGAAGACCATGATGGCATAGAA 654
|||||
Db 361 CAGAAGGATTTCTAGACGCTTATTACAACTCTGACCTGAAGACCATGATGGCATAGAA 420
|||||
QY 655 TGGGTCATCGCTTTTGTCTCAGCGGGGCTTTGTGATGAAACAGATCAGCATGTC 714
|||||
Db 421 TGGGTCATCGCTTTTGTCTCAGCGGGGCTTTGTGATGAAACAGATCAGCATGTC 480
|||||
QY 715 ATCAATGTTCACTATCTGACTGAACCTGCTTCTGAAGA 751
|||||
Db 481 ATCAATGTTCACTATCTGACTGAACCTGCTTCTGAAGA 517
|||||

RESULT 14
LOCUS AC020851 196900 bp DNA linear HTG 15-JUL-2000
DEFINITION Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
unordered pieces.
ACCESSION AC020851
VERSION AC020851.2 GI:9211212
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 15, 2000 this sequence version replaced gi:6696457.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1437240
Center clone name: RPCI-21_467L12
-----
Summary Statistics
Consensus quality: 164618 bases at least Q40
Consensus quality: 180867 bases at least Q30
Consensus quality: 183777 bases at least Q20
Estimated insert size: 147000; pulse field gel estimation
Estimated insert size: 193700; sum-of-contigs estimation
Quality coverage: 5.67 in Q20 bases; pulse field gel estimation
Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
* 1541: contig of 1541 bp in length
* 1542: gap of unknown length
* 1642: contig of 1851 bp in length
* 3493: gap of unknown length
* 3593: contig of 1146 bp in length
* 4739: gap of unknown length
* 4838: contig of 2410 bp in length
* 4839: gap of unknown length
* 7249: contig of 1830 bp in length
* 7349: gap of unknown length
* 9179: contig of 2215 bp in length
* 9279: gap of unknown length
* 11494: contig of 2014 bp in length
* 11594: gap of unknown length
* 13608: contig of 2239 bp in length
* 13708: gap of unknown length
* 15947: gap of unknown length
* 16047: contig of 2346 bp in length
* 18393: gap of unknown length
* 18493: contig of 2739 bp in length
* 21232: gap of unknown length
* 21332: contig of 3331 bp in length
* 24653: gap of unknown length
* 24763: contig of 2378 bp in length
* 27141: gap of unknown length
* 27241: contig of 2710 bp in length
* 29951: gap of unknown length
* 30051: contig of 2195 bp in length
* 32246: gap of unknown length
* 32346: contig of 3343 bp in length
* 35689: gap of unknown length
* 35789: contig of 4599 bp in length
* 40487: gap of unknown length
* 43580: contig of 3093 bp in length
* 43581: gap of unknown length
* 43681: contig of 2757 bp in length
* 46438: gap of unknown length
* 46538: contig of 3647 bp in length
* 50184: gap of unknown length
* 50285: contig of 4288 bp in length
* 54672: gap of unknown length
* 54673: contig of 5555 bp in length
* 60228: gap of unknown length
* 60328: contig of 7697 bp in length
* 68025: gap of unknown length
* 68125: contig of 10486 bp in length
* 78611: gap of unknown length
* 78711: contig of 9441 bp in length
* 88151: gap of unknown length
* 88251: contig of 8006 bp in length
* 96258: gap of unknown length
* 96358: contig of 9870 bp in length
* 106228: gap of unknown length
* 106328: contig of 9468 bp in length
* 115796: gap of unknown length
* 115896: contig of 10468 bp in length
* 126364: gap of unknown length
* 126464: contig of 10726 bp in length
* 137190: gap of unknown length
* 137290: contig of 9089 bp in length
* 146379: gap of unknown length
* 146479: contig of 15208 bp in length
* 161679: gap of unknown length
* 161787: contig of 14363 bp in length
* 176150: gap of unknown length
* 176250: contig of 20651 bp in length.
```

```

FEATURES
  source
    Location/Qualifiers
      1..196900
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="RP21-467L12"
        /clone_lib="RPGI mouse PAC library 21"
BASE COUNT      51456 a 43823 c 44608 g 53666 t 3347 others
ORIGIN
Query Match      37.9%; Score 284.4; DB 2; Length 196900;
Best Local Similarity 72.8%; Pred. No. 1.4e-70;
Matches 412; Conservative 0; Mismatches 141; Indels 13; Gaps 3;
QY 189 ATCTCATGCTTTTGAGGCTAATCATTTGGATTGTTCTCTTCAGATGGCTTTCCCGAA 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102413 ACCCCGACACTTCAGGGCTGCGCTTTGTCTTTGTT-CTTTCAGATGGCTCAATGAA 102471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 GATCAGATTCATGATATATTTGCTTCTGTTCTGTTGGGGCTCTTTGTTGTTATTTAGCAT 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102472 GACAAGGCTAGTTACGCTCCATCTCATGATGGCGCACCTGCTGTTGTTCTAGCAT 102531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 309 GTACAGTCTAAATCCCTTCAAGAACAGCTCTTTGTTTCAAGAAAG---ACGGAACTT 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102532 GG-----ATCTTTCAGAACTCCCGTTTGTGTTTAAAGAAAGTCACGGAACTT 102582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 CTTTAAGCTCCAGATACAGCTGACGAGCAGACACCTCCCTTCCTGCTCTCTGCTGAC 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102583 CCTTCAGATCCGATATACATGCAAGCAGAGCGCGCTTTCCTGGTCTCTGCTGAC 102642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 CTATCCCAACAAAGTTGGCTGAGCGCATGGCCATCCGCGACAGCTGGGGGAAAGAGAG 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102643 GTCATCTCAACAGCAGTGGCCCTCGCATGGCCATCCGCAAGCAGCTGGGGGTAGAGAG 102702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 486 GATGTCGAAGGAAGCAGCTGAAGACATTTCTTCCTCCCTGGGACCACACGAGTGCAGC 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102703 ATCTGTCAGGGCCACAGAGTGAGACCTTCTCTCTTGGGACCTCCGACAGACCGA 102762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 546 GGAACGAAGAGGTGGACAGAGAGCAGCAGCAGCAGCGGAGCATATTCAGAAAGATTT 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102763 GGAGATGACGCCACACACCTTGAGAGCAGCAGCAGCAGCAGCATATTCAGAAAGACTT 102822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 606 CCTAGAGCTCTATTACATCTGACCTGAAGACCATGATGGGCATAGATGGTTCATCG 665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102823 CAAGGATGCTACTTCACTGACCTGAGACCATGATGGGTATGGATGGGTCTACCA 102882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 666 CTTTGTCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAATGTTGA 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102883 CTTTGTCTCAGACAGCTTACCTGATGAAACAGGACTCTGACATGTTTGTGAATGTTGG 102942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 726 CTATCTGACTGAAGTCTTCTGAAGA 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102943 CTATCTGACGGAAGTCTGCTAAAGA 102968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AC120346/c 175861 bp DNA linear HTG 09-MAY-2002
LOCUS Mus musculus clone Rp23-147E11, WORKING DRAFT SEQUENCE, 22 ordered
DEFINITION pieces.
ACCESSION AC120346
VERSION AC120346.3 GI:20514894
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus.
ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 175861)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone Rp23-147E11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175861)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
  Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,

```

```

Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Nobu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodores,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 175861)
REFERENCE
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
  Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
  Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
  Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
  Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
  Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
  Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
  Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
  Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
  Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
  Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
  McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
  Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
  Nobu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
  Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
  Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
  Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
  Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodores,J.,
  Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
  Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
  Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 9, 2002 this sequence version replaced gi:20503171.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20921
Center clone name: 147_E11
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 169879 bases at least Q40
Consensus quality: 172188 bases at least Q30
Consensus quality: 173133 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 173761; sum-of-contigs
Quality coverage: 8.3 in Q40 bases; agarose-fp
Quality coverage: 8.4 in Q20 bases; sum-of-contigs
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 28835: contig of 28835 bp in length

* 28836 28935: gap of 100 bp

* 28936 29562: contig of 627 bp in length

* 29563 29662: gap of 100 bp

* 29663 30328: contig of 666 bp in length

* 30329 30428: gap of 100 bp

* 30429 32005: contig of 1577 bp in length

* 32006 32108: gap of 100 bp

* 32109 33128: contig of 1023 bp in length

* 33129 33228: gap of 100 bp

* 33229 34512: contig of 1284 bp in length

* 34513 34612: gap of 100 bp

* 34613 36079: contig of 1467 bp in length

* 36080 36179: gap of 100 bp

* 36180 37483: contig of 1304 bp in length

* 37484 37583: gap of 100 bp

* 37584 39166: contig of 1583 bp in length

* 39167 39266: gap of 100 bp

* 39267 41065: contig of 1799 bp in length

* 41066 41165: gap of 100 bp

* 41166 43203: contig of 2038 bp in length

* 43204 43303: gap of 100 bp

* 43304 47489: contig of 4186 bp in length

* 47490 47589: gap of 100 bp

* 47590 50604: contig of 3015 bp in length

* 50605 50704: gap of 100 bp

* 50705 56370: contig of 5666 bp in length

* 56371 56470: gap of 100 bp

* 56471 64079: contig of 7609 bp in length

* 64080 64179: gap of 100 bp

* 64180 72619: contig of 8440 bp in length

* 72620 72719: gap of 100 bp

* 72720 81566: contig of 8847 bp in length

* 81567 81666: gap of 100 bp

* 81667 96774: contig of 15108 bp in length

* 96775 96874: gap of 100 bp

* 96875 119647: contig of 22773 bp in length

* 119648 119747: gap of 100 bp

* 119748 147772: contig of 28025 bp in length

* 147773 147872: gap of 100 bp

* 147873 174557: contig of 26685 bp in length

* 174558 174657: gap of 100 bp

* 174658 175861: contig of 1204 bp in length.

FEATURES

Source

1..175861
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-147E11"
/clone_lib="RPCI-23 Female Mouse BAC"

misc_feature

1..28835

/note="assembly_fragment"
clone_end:SP6
vector_slide:left

misc_feature

28936..29562

/note="assembly_fragment"

misc_feature

29663..30328

/note="assembly_fragment"

misc_feature

30429..32005

/note="assembly_fragment"

misc_feature

32106..33128

/note="assembly_fragment"

misc_feature

33229..34512

/note="assembly_fragment"

misc_feature

34613..36079

misc_feature /note="assembly_fragment"
36180..37483
misc_feature /note="assembly_fragment"
37584..39166
misc_feature /note="assembly_fragment"
39267..41065
misc_feature /note="assembly_fragment"
41166..43203
misc_feature /note="assembly_fragment"
43304..47489
misc_feature /note="assembly_fragment"
47590..50604
misc_feature /note="assembly_fragment"
50705..56370
misc_feature /note="assembly_fragment"
56471..64079
misc_feature /note="assembly_fragment"
64180..72619
misc_feature /note="assembly_fragment"
72720..81566
misc_feature /note="assembly_fragment"
81667..96774
misc_feature /note="assembly_fragment"
96875..119647
misc_feature /note="assembly_fragment"
119748..147772
misc_feature /note="assembly_fragment"
147873..174557
misc_feature /note="assembly_fragment"
174558..175861
misc_feature /note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 48903 a 39540 c 39029 g 46286 t 2103 others
ORIGIN

Query Match 37.7%; Score 282.8; DB 2; Length 175861;

Best Local Similarity 72.6%; Pred. No. 4e-70;

Matches 411; Conservative 0; Mismatches 142; Indels 13; Gaps 3;

QY 189 ATCCTCATGCTTTGAGGTCTAATCATTTGATTTTTCCTTTTCAGATGGTTCCTCCGAA 248

Db 62918 ACCCGACACTTCTCAGCGCTGGCGTTTGTCTTTGTT-CITTCAGATGGTTCACATGAA 62860

QY 249 GATGAGATTGATGATATTTGCTTCTGGTCTGGGGCTCTTTGTTGTTATTAGCAT 308

Db 62859 GACAAGGCTAGTTTACGGCTCCATTCTCATGATGGCGCACTCTGCTGTACTTCAGCAT 62800

QY 309 GTACAGTCTAAATCCTTTCAAAGAACAGTCTCTTTTACAGAAAG---ACGGGAAGTT 365

Db 62799 GG-----ATTCTTCAGAGAACTCCCGTTTGTGTTTAAAGAAAGTCACGGGAAGTT 62749

QY 366 CCTTAGCTCCAGATACAGCTGCAGGCAGACACCTCCCTTCCTGCTCTGCTGGTGAC 425

Db 62748 CCTTCAGATTCGGATATAGACTGCAAGCAAGCGCCCTTTCCTGCTGCTGCTGGTGAC 62689

QY 426 CTCATCCCAACAAAGTTGGCTGAGCGCATGGCCATCCGCGCAGAGCTGGGGAAAGAGAG 485

Db 62688 GTCACTCAGAGAGAGCTGGCGCTCGCATGCCATCCGCAAGAGCTGGGTAGAGAGAC 62629

QY 486 GATGGTGAAGGAAAGCAGCTGAAGACATTTCTTCTCTGGGGACCCACGACGTGACG 545

Db 62628 ATCTGTGAGGGCCAAACAGGTGAGGACCTTCTTCTTCTGGGGACCCGACGACCGGA 62569

QY 546 GGAAGCAAGAGAGTTGGACAGGAGGAGCAGCAGCAGCGGACATATCCAGAGAGATT 605

Db 62568 GAGATGGAGCCACAAACCTGGAGAGGAGCAGCAGCAGCCGACATATCCAGAGAGATT 62509

QY 606 CCTAGAGCTCTATTACATCTGACCTGAAGACCATGATGGGCATAGATGGGTCCATCG 665

Db 62508 CAAGGATGCTACTTCAACCTGACCTGAAACCATGATGGGTATGGAATGGTCTACCA 62449

QY 666 CTTTGTCTCAGCGGGCTTTGTGATGAAACAGACTCAGACATGTTTCATCATGTTGA 725

|||||
Db 62448 CTTTGTCTCAGACAGCTTACGTGATGAAACGGACTCTGACATGTTTGTGAATGTTGG 62389

Qy 726 CTATCTGACTGAAGTCTTCTGAAGA 751

Db 62388 CTATCTGACGGAACTGCTGCTAAAGA 62363

Search completed: April 11, 2003, 21:15:33
Job time : 1837.46 secs

GenCore version 5.1.1.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:50:38 ; Search time 124.796 Seconds
(without alignments)
13552.085 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750
Perfect score: 751
Sequence: 1 ggaccatgccagaatctc.....gactgaactgtcttgaaga 751

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*

24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	751	100.0	10562	21	AAA93876 Human beta3Gal-T5
2	590.2	78.6	1011	21	AAA27959 Human beta3Gal-T5
3	517.4	68.9	2775	21	AAA93875 Human beta-1,3 gal
4	119.2	15.9	1739	15	AAO67067 Beta-1,3-galactosyl
5	91	12.1	1773	20	AAK87193 Human Dendriac cDN
6	91	12.1	1897	24	AK51201 Human cDNA encodin
7	91	12.1	2095	21	AAZ65022 Membrane-bound pro
8	91	12.1	2095	22	AAAS46005 Human DNA encoding
9	91	12.1	2095	22	AAF92075 Human PRO1074 cDNA

10	91	12.1	2095	22	AAF44168 Human PRO1074 (UNC
11	91	12.1	2168	22	AAH15711 Human cDNA sequenc
12	91	12.1	2189	20	AAK87192 Human Dendriac cDN
13	81.6	10.9	1134	21	AAA58791 DNA encoding the b
14	81.2	10.8	1725	23	ABLI4319 Drosophila melanog
15	81.2	10.8	4218	23	ABLI4318 Drosophila melanog
16	76.6	10.2	1134	22	AAO93556 Human beta-1,3-gal
17	75.8	10.1	3033	23	ABLI4314 Drosophila melanog
18	75.8	10.1	3330	23	ABLI4322 Drosophila melanog
19	73.8	9.8	1266	20	RAK35710 cDNA encoding a pr
20	73.8	9.8	2420	20	AAK35711 haematopoietic ste
21	69	9.2	400	21	AAZ94080 Human prostate exp
22	67	8.9	583	23	ABV08134 Human prostate exp
23	67	8.9	741	23	ABV38048 Human secreted pro
24	67	8.9	1271	20	AAK87194 Human Brainiac-3 c
25	67	8.9	1271	20	AAK87194 DNA sequence of gl
26	66.2	8.8	1613	20	AAK26195 Human prostate exp
27	64.2	8.5	748	23	ABV25215 Human beta 1,3-N-a
28	63.6	8.5	1296	22	AAF29258 DNA encoding the b
29	63.6	8.5	1420	21	AAAS8790 Human PRO4344 nucl
30	63.6	8.5	1642	21	AAAS8115 cDNA encoding huma
31	63.6	8.5	1643	24	ABK69978 Drosophila melanog
32	63	8.4	1775	23	ABLI7703 Drosophila melanog
33	63	8.4	12427	23	ABLI7704 Human epidermold c
34	60	8.0	1116	19	AAV49599 Human beta1,3-N-ac
35	60	8.0	1208	24	AAK16945 Human beta 1,3-N-a
36	60	8.0	2180	22	AAF29257 Human epidermold c
37	60	8.0	2186	19	AAV49598 Nucleotide sequenc
38	60	8.0	2186	24	ABLI41994 Human cDNA encodin
39	60	8.0	2198	22	AAF93831 Human beta 1,3-N-a
40	60	8.0	2205	22	AAF29256 Human DNA encoding
41	60	8.0	2210	22	AAAS46068 Human prostate exp
42	60	8.0	2427	23	ABV22526 Human prostate exp
43	60	8.0	2427	23	ABV28341 Murine beta-1,3-ga
44	56.4	7.5	1167	21	AAAS3201 Human galactosyltr
45	54.4	7.2	1434	20	AAZ11234

ALIGNMENTS

RESULT 1
AAA93876
ID AAA93876 standard; DNA; 10562 BP.
XX
AC AAA93876;
XX
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta3Gal-T5 encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; beta3Gal-T5; ds.
XX
OS Homo sapiens.
XX
FN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
XX WPI; 2000-549409/50.
XX
XX Beta-1,3 galactose transferase and DNA encoding it, useful for
XX synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
XX digestive system cancer

XX Claim 31; Page 103-111; 123pp; Japanese.

PS This invention relates to a polypeptide (I) with beta-1,3 galactose

CC transferase activity, or variants of (I) comprising amino acid additions,

CC deletions and/or substitutions. Included in the invention is DNA encoding

CC all or part of (I); expression vectors containing the DNA, host cells

CC transformed by the vectors; a method for the preparation of the

CC polypeptide by culture of the transformants or by expression in the milk

CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3

CC galactose transferase protein transfers galactose by beta-1,3 bonding to

CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as

CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and

CC DNA encoding it are useful for the treatment and diagnosis of cancer of

CC the digestive system. The present sequence represents a Beta3gal-T5

CC encoding DNA sequence.

XX

SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 100.0%; Score 751; DB 21; Length 10562;

Best Local Similarity 100.0%; Pred. No. 7.2e-227;

Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCAATGCCCAAGTCTCGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGC 60

DB 8000 GGACCAATGCCCAAGTCTCGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGC 8059

QY 61 AGATCTTAGACCTTTGTGCTTAACTGTTTAAACACACAGACACCGAGCTTCTGTATGCAGC 120

DB 8060 AGATCTTAGACCTTTGTGCTTAACTGTTTAAACACACAGACACCGAGCTTCTGTATGCAGC 8119

QY 121 GAGGTTCTAGAGTTTCCAAACACGGGTCTCCTCTCCACCTCAGCTCTAGCATATAAA 180

DB 8120 GAGGTTCTAGAGTTTCCAAACACGGGTCTCCTCTCCACCTCAGCTCTAGCATATAAA 8179

QY 181 CTAGACATCCTCATGCTTTGAGGCTTAATCATTTGGATTTTGTCTTTCAGATGGCT 240

DB 8180 CTAGACATCCTCATGCTTTGAGGCTTAATCATTTGGATTTTGTCTTTCAGATGGCT 8239

QY 241 TTCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGCGGGCTCTTTGTTGTAT 300

DB 8240 TTCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGCGGGCTCTTTGTTGTAT 8299

QY 301 TTTAGCATGTACAGTCTAAATCCTTTCAAAGACAGTCTTTGTTTACAAGAACGGG 360

DB 8300 TTTAGCATGTACAGTCTAAATCCTTTCAAAGACAGTCTTTGTTTACAAGAACGGG 8359

QY 361 AACTTCCTTAAGTCCCAAGATACAGACTGCAGGACAGACCTCCTTCCTCGTCTGCTG 420

DB 8360 AACTTCCTTAAGTCCCAAGATACAGACTGCAGGACAGACCTCCTTCCTCGTCTGCTG 8419

QY 421 GTGACCTCATCCCAACAGTGGTGAGCGCATGGCCATCCGGCAGAGCTGGGGAAA 480

DB 8420 GTGACCTCATCCCAACAGTGGTGAGCGCATGGCCATCCGGCAGAGCTGGGGAAA 8479

QY 481 GAGAGATGGTGAAGGGAAGCAGCTGAAGACATCTTCTCTCTGGGGACACACAGCAGT 540

DB 8480 GAGAGATGGTGAAGGGAAGCAGCTGAAGACATCTTCTCTCTGGGGACACACAGCAGT 8539

QY 541 GCAGCGGAACCAAGAGAGTGGACAGGAGACAGGACAGCGGGGACATATCCAGAAG 600

DB 8540 GCAGCGGAACCAAGAGAGTGGACAGGAGACAGGAGACAGCGGGGACATATCCAGAAG 8599

QY 601 GATTTCTTAGAGCTTATTACAATCTGACCTGAAGACCATGATGGGCGATAGATGGTC 660

DB 8600 GATTTCTTAGAGCTTATTACAATCTGACCTGAAGACCATGATGGGCGATAGATGGTC 8659

QY 661 CATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAAT 720

DB 8660 CATCGCTTTTGTCTCAGCGCGGTTTGTGATGAAGAAACAGACTCAGACATGTTTCATCAAT 8719

QY 721 GTTGACTATCTGACTGAAGTCTCTCTGAAGA 751

XX

DB 8720 GTTGACTATCTGACTGAAGTCTCTCTGAAGA 8750

RESULT 2

AAA27959

ID AAA27959 standard; DNA; 1011 BP.

XX

AC AAA27959;

XX

DT 15-AUG-2000 (first entry)

XX

DE Human beta3Gal-T5 gene sequence.

XX

KW UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;

KW Beta3Gal-T5; ss: human; chromosome 21q22.3; galactosylation;

KW beta1,3-galactosyl glycosylated saccharide production; glycopeptide;

KW glycoprotein.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT CDS 79..1011

FT /**tag= a

FT /product= "Beta3Gal-T5"

FT primer_bind complement {79..98}

FT /**tag= b

FT primer_bind complement {150..170}

FT /**tag= c

FT primer_bind 991..1011

FT /**tag= d

XX

PN WO200029558-A1.

XX

PD 25-MAY-2000.

XX

PF 11-NOV-1999; 99WO-US26807.

XX

PR 13-NOV-1998; 98DK-0001483.

XX

PA (CLAU/) CLAUSEN H.

XX

PI Clausen H, Amado M;

XX

DR WPI: 2000-399728/34.

DR P-PSDB; AAY94641.

XX

PT Novel nucleic acid sequence encoding human

PT UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase

PT useful for obtaining beta 1,3-galactosyl glycosylated saccharides and

PT glycopeptides or glycoproteins

XX

PS Claim 7; Fig 1; 74pp; English.

XX

CC The present invention relates to a nucleic acid sequence encoding

CC UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase

CC (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at

CC carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence

CC represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is

CC located on human chromosome 21q22.3. Beta3Gal-T5 is a type II

CC transmembrane glycoprotein. The invention also relates to the beta3Gal-T5

CC protein sequence, a nucleic acid vector comprising the beta3Gal-T5

CC nucleotide sequence, a host cell comprising the vector, and a method for

CC the production of the beta3Gal-T5 protein from the host cells. The

CC methods of the invention can be used for recombinant production of

CC beta3Gal-T5 for use as a catalyst and for recombinant production of

CC peptides or proteins with appropriate galactosylation. The beta3Gal-T5

CC protein can be used to obtain beta1,3-galactosyl glycosylated

CC saccharides, glycopeptides or glycoproteins.

XX

SQ Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 other;

Query Match 78.6%; Score 590.2; DB 21; Length 1011;

Best Local Similarity 99.5%; Pred. No. 1.9e-176;

Matches 592; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY 157	CCACCTCAGCCTCCCTAGCATAAACTAGACACATCCCTCATGCTTTTGGAGTCTAATCATTT 216
Db 1	CCACCTCAGCCTCCCTAGCATAAACTAGACACATCCCTCATGCTTTTGGAGTCTAATCATTT 60
QY 217	GGATTGTTTCTCTTTCAGATGGCTTTCCGGAAGATGAGATGATGATATTTGGCTTCTG 276
Db 61	GGATTGTTGTTCTCTTTCAGATGGCTTTCCGGAAGATGAGATGATGATATTTGGCTTCTG 120
QY 277	GTTCCTGGGGCTCTTTGTTTATTTAGCATGTACAGTCTAAATCCTTTCAAGAAGACAG 336
Db 121	GTTCCTGGGGCTCTTTGTTTATTTAGCATGTACAGTCTAAATCCTTTCAAGAAGACAG 180
QY 337	TCCTTTGTTTACAAAGAGAGGGAACTTCCTTAAGTCCCAAGATCAGACTGCAGGCAG 396
Db 181	TCCTTTGTTTACAAAGAGAGGGAACTTCCTTAAGTCCCAAGATCAGACTGCAGGCAG 240
QY 397	ACACCTCCCTTCTCGTCTCTGCTGCTGACCTCATCCCAACACAGTTGGCTGAGCGCATG 456
Db 241	ACACCTCCCTTCTCGTCTCTGCTGACCTCATCCCAACACAGTTGGCTGAGCGCATG 300
QY 457	GCATCCCGCAGACGTGGGGAAGAGAGATGGTGAAGGAAAGCAGCTGAAGACATTC 516
Db 301	GCATCCCGCAGACGTGGGGAAGAGAGAGATGGTGAAGGAAAGCAGCTGAAGACATTC 360
QY 517	TTCCCTCTGGGACACACAGCAGTGCAGCGGAAGAGAGGTGGACAGGAGAGCCAG 576
Db 361	TTCCCTCTGGGACACACAGCAGTGCAGCGGAAGAGAGGTGGACAGGAGAGCCAG 420
QY 577	CGACACGGGGACATTTATCCAGAAGGATTTCCCTAGACGCTCTATTACAATCTGACCCCTGAAG 636
Db 421	CGACACGGGGACATTTATCCAGAAGGATTTCCCTAGACGCTCTATTACAATCTGACCCCTGAAG 480
QY 637	ACCATGATGGGATAGATGGTCCATCGCTTTGCTCAGCGCGGCTTTGTCATGAAA 696
Db 481	ACCATGATGGGATAGATGGTCCATCGCTTTGCTCAGCGCGGCTTTGTCATGAAA 540
QY 697	ACAGACTCAGACATGTTTCATCAATTTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 541	ACAGACTCAGACATGTTTCATCAATTTTGACTATCTGACTGAACCTGCTTCTGAAGA 595
RESULT 3	
AAA93875	
ID	AAA93875 standard; DNA; 2775 BP.
XX	AAA93875;
AC	AAA93875;
DT	15-JAN-2001 (first entry)
XX	Human beta-1,3 galactose transferase encoding DNA.
XX	Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW	digestive system; ds.
OS	Homo sapiens.
XX	WO200050608-A1.
PN	31-AUG-2000.
XX	24-FEB-2000; 2000WO-JP01070.
PF	25-FEB-1999; 95JP-0047571.
PR	(KYOW) KYOWA HAKKO KOGYO KK.
XX	Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
PI	WPI: 2000-549409/50.
DR	P-PSDB; AAB93875.
XX	

PT	Beta-1,3 galactose transferase and DNA encoding it, useful for
PT	synthesis of type 1 sialyl Lewis ^x , a carbohydrate for treatment of
XX	digestive system cancer
PS	Claim 5; Page 99-102; 123pp; Japanese.
XX	This invention relates to a polypeptide (I) with beta-1,3 galactose
CC	transferase activity, or variants of (I) comprising amino acid additions,
CC	deletions and/or substitutions. Included in the invention is DNA encoding
CC	all or part of (I); expression vectors containing the DNA, host cells
CC	transformed by the vectors; a method for the preparation of the
CC	polypeptide by culture of the transformants or by expression in the milk
CC	of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC	galactose transferase protein transfers galactose by beta-1,3 bonding to
CC	N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC	GlcNAc6Sial-3galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC	DNA encoding it are useful for the treatment and diagnosis of cancer of
CC	the digestive system. The present sequence represents Beta-1,3 galactose
CC	transferase encoding DNA.
XX	Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 other;
SQ	
Query Match 68.9%; Score 517.4; DB 21; Length 2775;	
Best Local Similarity 99.8%; Pred. No. 4.2e-153;	
Matches 518; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 233	AGATGGCTTTCCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGGGGCTCTTT 292
Db 400	AAATGGCTTTCCGGAAGATGAGATGATGATATTTGCTTCTGCTTCTGGGGCTCTTT 459
QY 293	GTTCGTATTTTACATGTACAGTCTAAATCCTTTCAAGAAGACAGTCTTTTACAGA 352
Db 460	GTTCGTATTTTACATGTACAGTCTAAATCCTTTCAAGAAGACAGTCTTTTACAGA 519
QY 353	AGACGGGAACCTTCCTTAAGTCCCAAGATACAGACTGCAGGACAGACCTCCCTTCCTCG 412
Db 520	AGACGGGAACCTTCCTTAAGTCCCAAGATACAGACTGCAGGACAGACCTCCCTTCCTCG 579
QY 413	TCTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCGGCAGAGCT 472
Db 580	TCTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCGGCAGAGCT 639
QY 473	GGGGAAAGAGAGGATGGTGAAGGAAAGCAGCTGAAGACATCTCTCTCTGGGGACCA 532
Db 640	GGGGAAAGAGAGGATGGTGAAGGAAAGCAGCTGAAGACATCTCTCTCTGGGGACCA 699
QY 533	CCAGCAGTGCAGCGGAAGAGAGGTGGACAGGAGCGAGCGACACACGGGACATTA 592
Db 700	CCAGCAGTGCAGCGGAAGAGAGGTGGACAGGAGCGAGCGACACACGGGACATTA 759
QY 593	TCAGAAAGATTTCTTAGACGCTATTACATCTGACCTGAAGACCATGATGGGCATAG 652
Db 760	TCAGAAAGATTTCTTAGACGCTATTACATCTGACCTGAAGACCATGATGGGCATAG 819
QY 653	AATGGTCCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAAAACAGACTCAGACATGT 712
Db 820	AATGGTCCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAAAACAGACTCAGACATGT 879
QY 713	TCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 880	TCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAAGA 918
RESULT 4	
AAQ67067	
ID	AAQ67067 standard; cDNA to mRNA; 1739 BP.
XX	AAQ67067;
AC	AAQ67067;
XX	14-MAR-1995 (first entry)
DT	Beta-1,3-galactosyltransferase cDNA.
XX	

```
KW Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pAMOPRWML;
KW KJM-1 cells; ds.
XX Homo sapiens.
XX
XX Location/Qualifiers
FH Key 676..1656
FT CDS /tag= a
FT /product= Beta-1,3-galactosyltransferase
XX
XX JP06181759-A.
XX
XX 05-JUL-1994.
XX
XX 16-DEC-1992; 92JP-0336436.
XX
XX 16-DEC-1992; 92JP-0336436.
XX
XX (YOW ) KYOWA HAKKO KOGYO KK.
XX
XX WPI; 1994-251683/31.
XX
XX P-PSDB; AAR57433.
XX
XX Beta-galactosyl-transferase DNA and protein - useful for prodn. of
XX saccharide chains
XX
XX Claim 3; Page 22-24; 47pp; Japanese.
XX
XX This sequence encodes a beta-1,3-galactosyltransferase. This enzyme
XX can be used to produce physiologically active saccharide chains and
XX variants, and for improvement of saccharide chains bound to
XX physiologically active proteins. This cDNA represents a fragment
XX of the plasmid pAMOPRWML which was cloned in KJM-1 cells.
XX
XX Sequence 1739 BP; 473 A; 407 C; 422 G; 437 T; 0 other;
XX
XX Query Match 15.9%; Score 119.2; DB 15; Length 1739;
XX Best Local Similarity 56.7%; Pred. No. 6.8e-27;
XX Matches 220; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
XX
Qy 362 ACTTCCTTAAGCTCCAGATACAGACTGAGCGAGACACCTCCCTTCCTCTCTCTCTCGG 421
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 866 AATTTCCTTATCAAGGAGCCCAATAAATGTCAGAAACAACTTCCTTTCTTTATCTCA 925
Qy 422 TGACCTATCCCAACAGTTGGCTGAGCGCATGGCCATCCGCGAGAGCTGGGGAAAG 481
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 926 TCAGCCCACTCAAGGAATTGATCCCGTCAGGCAATCAGAGAGAGCTGGGGGATG 985
Qy 482 AGAGGTGGTGAAGGAAAGCAGCTGAAGACATTCTTCCTCTGGGGACCACAGCAGTG 541
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 986 AGAACAACTTTAAGGGATCAAGATACCCACCTGTCCTCTGGGCAAGATGCTGATC 1045
Qy 542 CAGCGGAACCAAGAGGTGGACAGAGAGCCAGCGACACGGGGACATTTCCAGAGG 601
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 1046 CTGTCTCAATCAGATGGTGGAGCAAGAGAGCCAAATCTTCATGATATCATCTGGAG 1105
Qy 602 ATTCCTAGAGCTGTTATTAATCTGACCTGAAGACATGATGGGCAAGATGGGTCC 661
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 1106 ACTTATGATCTCTACCAATACCTTACCTCAAACTTAATGGGATGAGATGGTGG 1165
Qy 662 ATCGCTTTTGTCTCAGCGGCGTTTGTGATGAACAGACTCAGACATGTTTCATCAATG 721
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 1166 CCACCTTTTGTTCAAAAGCAAGTATGTCATGAAACAGACAGCGACATTTTGTAAACA 1225
Qy 722 TTGACTATCTGACTGAAGCTGTTCTGAA 749
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 1226 TGGACAATCTTATTATTAATTAATCTGAA 1253
XX
XX RESULT 5
XX AAX87193
XX ID AAX87193 standard; cDNA; 1773 BP.
XX
```

```
AC AAX87193;
XX
XX 27-SEP-1999 (first entry)
XX
XX Human Dendriac cDNA.
XX
XX Dendriac; Brainiac-2; human; Notch; immune disorder;
XX neurological disorder; diagnosis; therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 20..980
FT /*tag= a
FT sig_peptide 20..94
FT /*tag= b
FT mat_peptide 95..977
FT /*tag= c
XX
XX WO9931116-A1.
XX
XX 24-JUN-1999.
XX
XX 17-DEC-1998; 98WO-US27049.
XX
XX 17-NOV-1998; 98US-0108928.
XX
XX 18-DEC-1997; 97US-0068006.
XX
XX 12-MAR-1998; 98US-0077687.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ebner R, Endress GA, Florence KA, Rosen CA, Ruben SM;
XX Soppet DR, Yu G;
XX
XX WPI; 1999-430031/36.
XX
XX P-PSDB; AAY06461.
XX
XX Human proteins for treating and detecting immune and neurological
XX disorders
XX
XX Claim 2; Page 119-120; 132pp; English.
XX
XX This is the nucleotide sequence of a cDNA clone which codes for
XX human Dendriac (see AAY06461), also called Brainiac-2, a novel
XX member of the Brainiac family. Dendriac cDNA (see also AAX87193) is
XX deposited as ATCC 203056 and ATCC 209627. It was initially
XX discovered in a dendritic cell cDNA library. Additional clones of
XX the same gene were identified in NTERA2 cell, adult pulmonary
XX tissue, salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth
XX muscle, cerebellum, 8-week-old human mbyo, haemagopoietoma,
XX amygdala, substantia nigra and whole brain cDNA libraries. Dendriac
XX nucleic acids are used in the recombinant production of Dendriac
XX polypeptides, especially the mature protein or epitope-bearing
XX fragments. Dendriac and Brainiac proteins are members of the Notch
XX family of proteins involved in the control of cell division. The
XX proteins may be used to detect and/or treat immune system and
XX neurological disorders.
XX
XX Sequence 1773 BP; 568 A; 300 C; 339 G; 566 T; 0 other;
XX
```

Query Match 12.1%; Score 91; DB 20; Length 1773;
Best Local Similarity 53.3%; Pred. No. 6e-18;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;

```
Qy 344 TTTACAGAAAGACGGGAACCTTAAAGTCCAGATACAGACTGCGAGCAGACCTC 403
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 157 TTTACAGAAAGACTTTCACCTTCACACTTCGAGAGCAATTCAAACTGCTCTCAAAATC 216
Qy 404 CCTTCCTCGTCTGCTGACCTCATCCCAACAGTTGGCTGAGCGATGGCCATCC 463
Dy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 217 CATTTCTGGTCACTTCTGGTGACCTCCACCTTCAGATGTGAAAGCCAGCGCCATTA 276
Qy 464 GCGACGCTGGGGAAAGAGAGAGGATGTTGAGGGGAAAGAGAGCTGAGACATTTCTCTCC 523
```

```

Db 277 GAGTTACTTGGGGTGAATAAAGTCTTGTTGGGGATATCAGGTTCTTACATTTTCTTAT 336
QY 524 TGGGGACCACAGCAGTGCAGCGGAAACGAAAGAGG-----TGGACACGAGAGACC 574
Db 337 TAGGCCAAGAGGCTGAAAGGAAGACAAATGTTGGCATTGTCCTTAGAGGATGAACACC 396
QY 575 AGGACACGGGACATATTCAGAGGATTTCCTAGAGCTCTATTACAACTGACCCCTGA 634
Db 397 TTCTTTATGTTGACATAATCCGACAAAGATTTTTATAGACACATATAATAACCTGACCTTGA 456
QY 635 AGACCATGATGGCATAGAAATGGGTCATCGCTTTTGTCTCAGCGCGGTTTGTGATGA 694
Db 457 AAACCAATTATGGCAATTCAGTGGTAACTGAGTCTTGGCCCAATGCCAAGTACGTAATGA 516
QY 695 AAACAGACTGACACATGTTTCATCAATGTTGACTATCTGACTGAAGTCTTCTTGA 749
Db 517 AGACAGACACTGATGTTTTCATCAATGTTGCAATTTAGTGAAGTATCTTTTAA 571

RESULT 6
ABK51201
ID ABK51201 standard; cDNA; 1897 BP.
XX AC
XX AC
XX AC
XX 30-JUL-2002 (first entry)
XX DE
XX DE Human cDNA encoding betal,3-acetylglalactosamine transferase.
XX KW
XX KW Human; ss; gene; betal,3-acetylglalactosamine transferase;
XX KW Gb4 sugar chain.
XX OS
XX OS Homo sapiens.
XX PH
XX PH Location/Qualifiers
FT misc_feature 109..1101
FT /tag= a
FT /note= "This region is specifically claimed in claim 1"
FT /note= "109..1104"
FT CDS 109..1104
FT /tag= b
FT /product= "Betal,3-acetylglalactosamine transferase"
FT polyA_signal 1157..1162
FT /tag= c

XX JP2002085069-A.
XX XX
XX 26-MAR-2002.
XX XX
XX 08-SEP-2000; 2000JP-0273835.
XX XX
XX 08-SEP-2000; 2000JP-0273835.
XX XX
XX (SEK ) SEIKAGAKU KOGYO CO LTD.
XX (FURU/) FURUKAWA K.
XX XX
XX WPI; 2002-378274/41.
XX P-PSDB; AAU80224.
XX XX
XX Preparation of beta-1,3-acetylglalactosamine transferase -
XX XX
XX Claim 1; Page 11-13; 15pp; Japanese.
XX XX
XX The invention relates to the preparation of beta-1,3-acetylglalactosamine
XX (a Gb4 sugar chain molecule) comprising introducing a DNA appearing as
XX CC ABK51201 (S1) or a DNA hybridising with S1 or a base sequence
XX CC complementary to the base sequence or part of these base sequences under
XX CC a stringent condition into a cell and growing the cell to express beta-1,
XX CC 3-acetylglalactosamine transferase and collecting it. Also included is the
XX CC preparation of Gb4 sugar chain comprising contacting a polypeptide
XX CC comprising the beta-1,3-acetylglalactosamine transferase protein appearing
XX CC as AAU80224 (A) or a polypeptide consisting of an amino acid sequence in
XX CC which at least one amino acid is replaced, deleted, inserted or

```

```

CC transferred in the amino acid sequence (A) and having enzymatic activity
CC transferring N-acetylglalactosamine residue from an N-acetylglalactosamine
CC donor to the C3 site of the galactose residue in Gb3 sugar chain which is
CC the receptor with an N-acetylglalactosamine donor and Gb3 sugar chain.
CC The present sequence is the cDNA encoding beta-1,3-acetylglalactosamine
CC transferase.
XX SQ
XX Sequence 1897 BP; 596 A; 327 C; 366 G; 608 T; 0 other;
Query Match 12.1%; Score 91; DB 24; Length 1897;
Best Local Similarity 53.3%; Pred. No. 6.3e-18;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;
QY 344 TTACAGAAGAAGACGGAACTTCCTTAAGCTCCCAATACAGACTGCAGGCAGACACCTC 403
Db 281 TTACAGACAAGACTTTCACCTTCGAGAGCAATTCACAACTGCTCTCATCAAAATC 340
QY 404 CCTTCCTCGCTCGTGGTGACCTCATCCACAAACAGTTGGTGAGCGCATGGCCATCC 463
Db 341 CATTCCTGTCATTCGTTGACCTCCACCTTCAGATGTGAAGCCAGGCGGCATTGA 400
QY 464 GGCAGAGCTGGGGAAAGAGAGGATGGTGAAGGGAAGAGCTGAAGACATCTTCCTCC 523
Db 401 GAGTTACTTGGGTGAAAAAAGTCTTGGTGGGATATGAGGTTCTTACATTTTCTTAT 460
QY 524 TGGGGACCCACGACGTCGACGGGAAACGAAAGAGG-----TGGACACGAGGAGCC 574
Db 461 TAGCCAAAGAGGCTGAAAGGAAGACAAATGTTGCAATGTCCTTAGAGGATGAACACC 520
QY 575 AGGCACACGGGACATATCCAGAAGGATTTCCTAGACGCTCTATTACAACTGACCCCTGA 634
Db 521 TTCTTTATGTTGACATAATCCGACAAAGATTTTTATAGACACATATAATAACCTGACCTTGA 580
QY 635 AGACCATGATGGCATAGAAATGGTCCATCGCTTTTGTCTCAGCGCGGTTTGTGATGA 694
Db 581 AAACCAATTATGGCAATTCAGTGGTAACTGAGTCTTGGCCCAATGCCAAGTACGTAATGA 640
QY 695 AAACAGACTGACACATGTTTCATCAATGTTGACTATCTGACTGAAGTCTTCTTGA 749
Db 641 AGACAGACACTGATGTTTTCATCAATGTTGCAATTTAGTGAAGTATCTTTTAA 695

RESULT 7
AAZ65022
ID AAZ65022 standard; cDNA; 2095 BP.
XX AC
XX AC AAZ65022;
XX DT
XX 05-APR-2000 (first entry)
XX XX
XX Membrane-bound protein PRO1074 encoding cDNA.
XX KW
XX KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
XX KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX OS
XX OS Homo sapiens.
XX XX
XX WO963088-A2.
XX XX
XX 09-DEC-1999.
XX XX
XX 02-JUN-1999; 99WO-US12252.
XX XX
XX 02-JUN-1998; 98US-0087607.
XX XX
XX 02-JUN-1998; 98US-0087609.
XX PR
XX PR 02-JUN-1998; 98US-0087759.
XX PR
XX PR 03-JUN-1998; 98US-0087827.
XX PR
XX PR 04-JUN-1998; 98US-0088021.
XX PR
XX PR 04-JUN-1998; 98US-0088025.
XX PR
XX PR 04-JUN-1998; 98US-0088028.
XX PR
XX PR 04-JUN-1998; 98US-0088029.
XX PR
XX PR 04-JUN-1998; 98US-0088030.
XX PR
XX PR 04-JUN-1998; 98US-0088033.

```

PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088736.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0088906.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091479.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091629.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 19-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 31-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX WPI: 2000-072883/06.
DR P-PSDB: AA66665.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS Claim 2; Fig 136; 822pp; English.
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block


```
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99US-0149396.
PR 13-SEP-1999; 99US-0149396.
PR 08-OCT-1999; 99US-0151547.
PR 30-NOV-1999; 99US-0158663.
PR 01-DEC-1999; 99US-0158663.
PR 16-DEC-1999; 99US-0158663.
PR 05-JAN-2000; 99US-0158663.
PR 06-JAN-2000; 99US-0158663.
PR 11-FEB-2000; 99US-0158663.
PR 18-FEB-2000; 99US-0158663.
PR 22-FEB-2000; 99US-0158663.
PR 24-FEB-2000; 99US-0158663.
PR 02-MAR-2000; 99US-0158663.
PR 15-MAR-2000; 99US-0158663.
PR 20-MAR-2000; 99US-0158663.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi C, Gurney AL, Kijavir IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
XX WPI: 2001-032160/04.
XX P-PSDB; AAB65208.
XX
XX PRO polynucleotides used to produce polypeptides used to target
XX bioactive molecules such as toxins, radiolabels or antibodies, to
XX specific cells, to cause targeted cell death -
XX
XX Claim 2; Fig 136; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
XX proteins. The PRO proteins have cytostatic activity. The PRO proteins
XX can be used for targeted delivery of bioactive molecules, such as
XX toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
XX sequences, and their fragments, can be used as hybridisation probes, in
XX chromosomal and gene mapping, and in the generation of anti-sense RNA
XX and DNA. They may also be used to produce transgenic animals which are
XX used to develop and screen therapeutically useful reagents. The PRO
XX nucleotide and protein sequence can be used for tissue typing and in
XX treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
XX AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
XX in the isolation of human PRO sequences. AAF44087 to AAF44269 and
XX AAB55134 to AAB55300 represent human PRO polynucleotide and protein
XX sequences given in the exemplification of the present invention.
XX
XX Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;
XX
XX Query Match 12.1%; Score 91; DB 22; Length 2095;
XX Best Local Similarity 53.3%; Pred. No. 6.6e-18;
XX Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;
XX
XX 344 TTACAGAAAGACGGGACCTTCCTTAAGCTCCAGATACAGATGACGAGGACGACCTC 403
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 494 TTACAGAAAGACCTTCCTTCACATTCGAGAGCATTCACAACTGCTCTCAAAATC 553
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 404 CCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 554 CAFTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 613
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 464 GCGACAGCTGGGGGAAAGAGAGGATGGTGAAGGAAAGACGCTGAAGACATTCCTCTCC 523
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 614 GAGTTACTTGGGTCGAAAGAGAGTCTTGGTGGGATATGAGGTTCTTACATTTTCTAT 673
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 524 TGGGACACACAGCAGCTGCGGCGGGAACGAAGAGG-----TGGACGAGGAGGCC 574
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


XX Human; beta-1,3-galactosyltransferase homologue; ZNSSP6;
KW glycoprotein synthesis; type II transmembrane protein; chromosome 12q24;
XX metastasis; ss.
XX Homo sapiens.
XX WO200041539-A2.
XX 20-JUL-2000.
XX 12-JAN-2000; 2000WO-US00748.
XX 12-JAN-1999; 99US-0229222.
XX (ZYMO) ZYMOGENETICS INC.
XX Conklin DC, Yamamoto G, Gao Z, Jaspers SR;
XX WPI; 2000-475909/41.
XX A new isolated polypeptide for the study, diagnosis, prevention and/or
PT treatment of metastases comprises a 256 amino acid sequence described
PT in the specification -
XX Disclosure: Page 101; 105pp; English.
XX The present sequence represents a degenerate sequence encoding the
CC beta-1,3-galactosyltransferase homologue ZNSSP6. The ZNSSP6
CC polypeptide and polynucleotide are cell-cell interaction and
CC glycoprotein synthesis modulating. The polypeptide is predicted
CC to be a type II transmembrane protein, and is expressed in the
CC brain, kidney, and testis. The ZNSSP6 gene is localised to
CC chromosome 12q24. The ZNSSP6 polypeptides and polynucleotides are
CC used for the study, diagnosis, prevention and/or treatment of
CC metastases.
XX SQ Sequence 1134 BP; 136 A; 139 C; 211 G; 159 T; 489 other;
Query Match 10.9%; Score 81.6; DB 21; Length 1134;
Best Local Similarity 33.7%; Pred. No. 4.5e-15;
Matches 138; Conservative 61; Mismatches 208; Indels 3; Gaps 1;
QY 337 TCCTTTGTTTACAGAAAGCGGGAACCTCTTAAGTCCAGATACAGACTGCAGCGAC 396
DB 286 TTYTNNACNTAYMGNCAYTGYMNAAYTYWSNATHYNTNGARCCNWSNGNTGYWSN 345
QY 397 ACACCTCCTCTCCTCCTCTGCTGACCTCATCCCAACACAGTTGGCTGAGCGCATG 456
DB 346 AARGAYACNTTYTNYNTNGCNATHAARWSNCARCCNGNCAYTNGARMNGMNGCN 405
QY 457 GCCATCCGGCAGACGTGGGGA---AAGAGAGATGGTGAAGGAAAGACGCTGAAGACA 513
DB 406 GCNATHMGNWSNACNTGGGNGMNGTNGNGGNTGGGNCMNGMNGNCARYTNAARYTN 465
QY 514 TTCCTCCTCTCTGGGACACACAGCAGTGCAGCGGAAAGAGAGTGGACGAGGAGC 573
DB 466 GTNTTYTNYNGNGTNGCNGWSNCCNCCNCCNARYTNYNGNTAYGARWSN 525
QY 574 CAGCGACAGCGGGAATTTATCCAGAAGATTTCTAGACGCTCTATTACAATCTGACCCCTG 633
DB 526 MGNARTTYGAYGAYATHTYNTCARTGGGAYTTACNGARGAYTTTYYAAYTYNACNYTN 585
QY 634 AAGACCATGAGGCGATAGATGGTCCATCGCTTTTGTCTCAGCGCGGCTTGTGATG 693
DB 586 AARGARYTNCAYTNYNCARMNGTGGTNGTNGCNGCNTGYCCNCCARGNCAYTAYTYTN 645
QY 694 AAAACAGACTCAGACATGTTTCAATGTTGACTATCTGACTGAACCTGCT 743
DB 646 AARGCNGAYGAYGTYTNTTYGTNCAYTNCNCAAYTNTNGARTTYT 695
RESULT 14

ABLL14319
ID ABL14319 standard; cDNA; 1725 BP.
XX ABL14319;
AC
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 37439.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PM
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PR
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-658860/75.
DR P-PSDB; ABB70216.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX Claim 1; SEQ ID NO 37439; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLL01840-ABLL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1725 BP; 481 A; 480 C; 433 G; 331 T; 0 other;
SQ
Query Match 10.8%; Score 81.2; DB 23; Length 1725;
Best Local Similarity 55.6%; Pred. No. 7.6e-15;
Matches 180; Conservative 0; Mismatches 138; Indels 6; Gaps 1;
QY 409 CTCGTCTCTGCTGCTACCTCATCCCAACACAGTTGGCTGAGCGATGGCCATCCGCGAC 468
DB 991 CTCCTGGTTCATTATTAGCTCAGCGATGCTCAGATGAGCCGCCATGTCATCCGCGAC 1050
QY 469 ACGTGGGGGAAAGAGAGGATGGTGAAGGAAAGAGAGTGAAGACATTTCTTCTCTGGGG 528
DB 1051 ACTTGGATGCATTA-----CGGGACGAGAGAGGAGCGTGGCATTTGTCTCTAGGA 1104
QY 529 ACCACAGCAGTGCAGCGGAAAGAGAGTGGACGAGAGAGCCAGCGACACGGGAC 588
DB 1105 CTGGCACCACATGAACCAACATAAACAAAGCGCTTACCCAGGAGACTTCATCTATGGAGAT 1164
QY 589 ATTATCCAGAAGGATTTCTAGACGCTATTACATCTGACCCCTGAAGACCATGATGGGC 648
DB 1165 CTGATACGGGGAACCTTCATGACTCGTACACACACCTCAGCTCAAGACAATATCGACG 1224
QY 649 ATAGAATGGTCCATCGCTTTTGTCTCAGCGGGGCTTTGTGTGATGAACACAGACTCAGAC 708
DB 1225 CTGGAATGGCAGATGTGCATTGGCCGAGGAGGAAAGTACATTCTCAAGACGCGAGCGAC 1284

QY 709 ATGTTTCATCAATGTTGACTATCTG 732
|||||
Db 1285 ATGTTTCATCAATGTTGCCAAGCTG 1308

RESULT 15

ABL14318
ID ABL14318 standard; CDNA; 4218 BP.
XX
AC ABL14318;
XX

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 37436.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

FA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

XX P-PSDB; ABB70215.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Claim 1; SEQ ID NO 37436; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4218 BP; 1228 A; 939 C; 917 G; 1134 T; 0 other;

Query Match 10.8%; Score 81.2; DB 23; Length 4218;
Best Local Similarity 55.6%; Pred. No. 1.3e-14;
Matches 180; Conservative 0; Mismatches 138; Indels 6; Gaps 1;

QY 409 CTCGTCTCTGCTGGTGAAGTCCACAAACAGTTGGCTGAGCGCATGGCCATCCCGGAG 468
|||||

Db 2484 CTCCTGGTCTTATTAGCTCAGCGATGCTCAGCATGACCCCGCATGTCATCCCGGAG 2543
|||||

QY 469 ACGTGGGGGAAGAGAGATGGTGAAGGGAAGCAGCGTGAAGACATCTTCCTCTGGGG 528
|||||

Db 2544 ACTTGGATGCATTA-----CGGGACGAGAAGGGACGTGGGCATTTGTCTTAGGA 2597
|||||

QY 529 ACCACACGACGTGACGGGAAACGAAGAGTGGACGAGGAGCCAGGCACACGGGGAC 588
|||||

Db 2598 CGTGGCCCAATGAACCATAAACAAGCGCTTACCAGGAGAACTTCATCTATGGAGAT 2657
|||||

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:35:54 ; Search time 836.08 Seconds
(without alignments)
14547.424 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750

Perfect score: 751

Sequence: 1 ggacaaatgccagaatctc.....gactgaactgtcttctgaaga 751

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pin:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_man:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	326	43.4	338	9	AJ003597
c 2	234	31.2	400	17	A2768940
c 3	162.4	21.6	636	13	B965086
c 4	132.6	17.7	1101	17	CNS03Brl
c 5	127	16.9	233	10	B959593
c 6	115.4	15.4	1021	17	CNS044T4
c 1	326	43.4	338	9	AJ003597
c 2	234	31.2	400	17	A2768940
c 3	162.4	21.6	636	13	B965086
c 4	132.6	17.7	1101	17	CNS03Brl
c 5	127	16.9	233	10	B959593
c 6	115.4	15.4	1021	17	CNS044T4

c 7	101.4	13.5	966	17	CNS04N8R
c 8	100.8	13.4	921	17	CNS03CTW
c 9	99.8	13.3	815	10	AW128625
c 10	95.6	12.7	611	13	BJ039496
c 11	93.8	12.5	793	12	BG207694
c 12	92.8	12.4	444	13	BM403751
c 13	91	12.1	605	14	BQ417019
c 14	89.4	11.9	3564	11	BC028571
c 15	87.8	11.7	1101	13	BM548138
c 16	87.4	11.6	583	13	B1441305
c 17	85.6	11.4	809	17	CNS04BGI
c 18	81.2	10.8	1514	11	AK003837
c 19	80	10.7	983	13	BM609433
c 20	76.2	10.1	596	13	BJ501651
c 21	73.8	9.8	965	17	CNS051B2
c 22	70.6	9.4	427	13	B1476269
c 23	70.4	9.4	1041	17	CNS03BMO
c 24	70	9.3	620	9	AU133606
c 25	69.2	9.2	581	17	AZ381343
c 26	68.4	9.1	879	14	BQ95238
c 27	68.2	9.1	712	9	AL042887
c 28	67.8	9.0	675	10	BB48067
c 29	67.8	9.0	1568	11	AK007600
c 30	66.2	8.8	675	12	BG077355
c 31	64	8.5	496	14	R15977
c 32	64	8.5	574	10	BE652985
c 33	61.8	8.2	825	9	AL555348
c 34	60.4	8.0	343	12	BF605662
c 35	59	7.9	486	14	R61672
c 36	59	7.9	884	9	AL522684
c 37	58.4	7.8	266	10	AW416413
c 38	58.4	7.8	538	10	BE682916
c 39	58.2	7.7	454	9	AA972783
c 40	57.4	7.6	546	10	BE237286
c 41	57.4	7.6	594	17	FR0013836
c 42	57.2	7.6	411	12	BE699313
c 43	57	7.6	192	14	W30001
c 44	56.8	7.6	752	10	BE307613
c 45	56.6	7.5	502	10	AW652526

ALIGNMENTS

RESULT 1.
AJ003597/c 338 bp mRNA linear EST 04-DEC-1997
LOCUS AJ003597 Selected chromosome 21 cDNA library Homo sapiens CDNA
DEFINITION AJ003597 Selected chromosome 21 cDNA library Homo sapiens CDNA
ACCESSION AJ003597
VERSION AJ003597.1 GI:2578270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 338)
AUTHORS Szulzewsky, I., Hunt, E., Nguyen, M., Korn, B., Roehrdanz, B., Lehrach, H. and Yaspo, M. L. H.
TITLE An integrated transcript map for the whole human chromosome 21
JOURNAL Unpublished (1997)
COMMENT Contact: yaspo, M. L.
Max Planck Institut fuer Molekulare Genetik
Inhestasse 73, D14195 Berlin-Bahlem, Germany.

FEATURES
Location/Qualifiers
1..338
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q"
/clone="MPI19-12J9"
/clone_lib="Selected chromosome 21 cDNA library"
/note="Swaroop et al. (1991) Nucleic Acids Res. 19: 1954."
BASE COUNT 76 a 92 c 87 t 1 others

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
1 (bases 1 to 1101)	
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.	
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
Unpublished	
2 (bases 1 to 1101)	
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.	
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
Unpublished	
3 (bases 1 to 1101)	
Unpublished	
Genoscope.	
Direct Submission	
Submitted (12-APR-2000)	
This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	

submitted (12 Apr 2007)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
 Location/Qualifiers
 1..1101
 FEATURES
 source

[illegible]


```

QY 446 CTGAGCGCATGGCCATCCGGCAGACAGTGGGGAAAGAGAGAGGATGGTGAAGGAAAGAGCAGC 505
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 305 ATCCCGTAACGCCATCCGCCACAGTGGGAAAGAGAGACAGCAGTGGGCGCTGGGCT 246
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 506 TGAAGACATCTCTCTCTGCTGGGACACACAGCAGTGCAGCGGAACGAAAGAGG----- 559
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 245 TCGTTCGCTCTCTCTGCTGGGAAAGAGGAGGAGTGCAGACCTTTCTCCAAAGCAGCA 186
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 560 TGGACACGAGAGAGCCAGCAGCAGCGGACATATCCAGAAGGATTTCTAGACGCTCTATT 619
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 185 TCAGGAAGAGAGAGCCAAATTTACCATGACATCATCCACAGGAGTACCAAGACACCTACT 126
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 620 ACAATCTGACCTGAGACCATGATGGCATAGATGGTTCATCGCTTTTCTCTCAGG 679
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 125 ACAACCTGACCAATAAACCCCTGATGGGTGATGAATGGTGGCCACCTATTGCCACACAG 66
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 680 CGCGTTTGTGTGTAAGAACAGACTCAGACATGTTTCATCAATGTTGACTATGACTGAA 738
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 65 CCTCCTCGTGTGTAAGAACAGACGCGACATGTTGTCAACACAGAGTATCHCAACCAA 7

RESULT 7
CNS04N8R/c 966 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 122A15 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL298404
VERSION AL298404.1 GI:8036984
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 966)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
UNPUBLISHED
JOURNAL 2 (bases 1 to 966)
REFERENCE Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
AUTHORS Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bertot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
UNPUBLISHED
JOURNAL 3 (bases 1 to 966)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000)
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..966
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="122A15"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG122AA08SP1-end :
PUC-Ori"
BASE COUNT 188 a 255 c 312 g 190 t 21 others
ORIGIN
Query Match 13.5%; Score 101.4; DB 17; Length 966;
Best Local Similarity 54.2%; Pred. No. 9.3e-18;
Matches 212; Conservative 9; Mismatches 161; Indels 9; Gaps 1;

```

```

QY 362 ACTTCCTTAAGCTCCCGAGATACAGACTGCAGGACGACACCTCCCTTCCTCTGCTGCTGG 421
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 705 ACTTGGTTATCAACACGCCCGCACAAATGGCAGCAGATGAAGCCCTTCCTGGTTCTGGTGG 646
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 422 TGACCTCATCCCAACAACAGTTGCTGAGCGCATGCCATCGGCGACAGTGGGGGAAAG 481
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 645 TACGGTGGCGCCCAACAACAARCTCTATCGGACGATATCCGCAAGACTCTGGGCAGCG 586
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 482 AGAGGATGTTGAAGGAAAGACAGCTGAAGACATTTCTTCCTCTGGGACACACAGCAGTG 541
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 585 AGAGCCCGCTACTGGGCGAGAGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 542 CAGCGG-----AAACGAAGAGGTGACACAGAGCAGCGACGACGCGGACATTA 592
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 525 AGCGCGCGAGGAGCTTCAGCAGCAGCTGTGGAGAGARCCGGGAGCATGGGACCTGG 466
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 593 TCCAGAAGGATTTCTCTAGACGCTCTATTACAATCTGACCTGAGACACATGATGGGCATAG 652
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 465 TCCAGGCGGACTTCTCTGAACCTCTCAAGAACCTGACCATCAAGACCATGTTGATGCTGG 406
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 653 AATGGGTCCATCGCTTTTGTCTCAGCGCGCTTTGTGATGAACACAGACTCAGACATGT 712
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 405 AGTGTGTCARGCTMACTCTCGGAGCTTCTTACGCCATGAARATCRAMTCGGACACT 346
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 713 TCATCAATGTTGACTATCTGACTGAACTGCT 743
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 345 TCTCAACGTGCCCAATCTCATCCGCATGCT 315
  || | ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

RESULT 8
CNS03CTW 921 bp DNA linear GSS 15-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 015M16 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL238253
VERSION AL238253.1 GI:7897388
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 921)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
UNPUBLISHED
JOURNAL 2 (bases 1 to 921)
REFERENCE Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
AUTHORS Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bertot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
UNPUBLISHED
JOURNAL 3 (bases 1 to 921)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000)
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..921
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="015M16"
/clone_lib="G"

```


LOCUS	BC028571	3564 bp	mRNA	linear	HTC 21-AUG-2002
DEFINITION	Homo sapiens, UDP-Gal-4-epimerase, clone IMAGE:4838965, mRNA.				
ACCESSION	BC028571				
VERSION	BC028571.1	GI:22382221			
KEYWORDS	HTC				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 3564)				
TITLE	Strausberg,R.				
JOURNAL	Direct Submission				
REMARK	Submitted (23-APR-2002) National Institutes of Health, Mammalian				
COMMENT	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
	NIH-MGC Project URL: http://mgc.nci.nih.gov				
	Contact: MGC help desk				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.				
	CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA sequencing by: Institute for Systems Biology				
	http://www.systemsbiology.org				
	contact: amadan@systemsbiology.org				
	Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov				
	Series: IRAK File: 34 Row: p Column: 23				
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 15451873				
	This clone has the following problem: incomplete processing.				
FEATURES	Location/Qualifiers				
source	1..3564				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4838965"				
	/tissue_type="Testis"				
	/clone_lib="NIH.MGC.97"				
	/lab_host="DH10B"				
	/note="Vector: pBluescript"				
BASE COUNT	1048 a	661 c	741 g	1114 t	
ORIGIN					
	Query Match 11.9%; Score 89.4; DB 11; Length 3564;				
	Best Local Similarity 53.0%; Pred. No. 4e-14;				
	Matches 220; Conservative 0; Mismatches 186; Indels 9; Gaps 1;				
Qy	344	TTTACAGAAACAGCGGAACTTCCTTAGCTCCACAGATACAGACTCGAGCAGACACCTC 403			
Db	864	TTTACAGACAGACITTCACITCACCTCGAGAGCATCAAACTGCTCATCAAAATC 923			
Qy	404	CCTTCCTCTGCTGTGTGACCTATCCCAACAACAGTTGGCTGAGCGATGGCGATCC 463			
Db	924	TATTTCTGCTCATCTTGTTGACCTCCCACTTCAGATGTGAAGACGAGCGACCATTA 983			
Qy	464	GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGAAAGCAGCTGAAGACATCTTCCTCC 523			
Db	984	GAGTTACTTGGGGTGAAAAAAGTCTTGGTGGGATATGAGTTCTTACATTTTCTTAT 1043			
Qy	524	TGGGGACACACAGCAGTGCAGCGGAAACGAAAGAGG-----TGGACCAAGAGACC 574			
Db	1044	TAGGCCAAGAGGCTGAAAGGAAGACAAATATTTGGCATTTGCTTAGAGGATGAACACC 1103			
Qy	575	AGCGACAGCGGACATATTCAGAGAGGATTTCTTAGACGCTTATTACATCTGACCTTGA 634			
Db	1104	TTCTTATGTTGACATATTCGACAAAGATTTTITTAGACACATATAATACCTGACCTTGA 1163			
Qy	635	AGACCATGATGGGCATAGAATGGGTGCCATCGCTTTTCTCGAGCGGGGTTTGTATGA 694			

Db 402 TAGGCCAGAGGCTGAAAGGAGACACAAATGTTGGCATTGCTTTACAGGATGAACACC 461
Qy 575 AGCGACACGGGGACATTATCCAGAAGGATTTCCCTAGACGTCCTATTACAACTCTGACCCCTGA 634
Db 462 TTCCTTTATGGTGACATAATCCGACAAGATTTTATAGACATATAATAAAGCTGACCTTGA 521
Qy 635 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGA 694
Db 522 CAACCATTTATGGCATTTCAGGTGGGTAACTGAGTTTTGCCCCCAATGCCAAGTACGTAATGA 581
Qy 695 AAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAAGTCTTCTGAA 749
Db 582 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAA 636

Search completed: April 11, 2003, 22:38:00
Job time : 841.08 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	81.6	10.9	1134	4	US-09-482-180A-3	Sequence 3, Appli	
	63.6	8.5	1420	4	US-09-482-180A-1	Sequence 1, Appli	
2	56.4	7.5	1167	4	US-09-459-133-14	Sequence 14, Appli	
3	54.4	7.2	1434	2	US-09-055-097-2	Sequence 2, Appli	
4	52.8	7.0	1191	4	US-09-459-133-3	Sequence 3, Appli	
5	41.6	5.5	1532	4	US-09-459-133-1	Sequence 1, Appli	
6	40.2	5.4	1464	4	US-09-459-133-12	Sequence 12, Appli	
7	40.2	5.4	7218	1	US-08-232-463-14	Sequence 14, Appli	
8	39.6	5.3	7218	1	US-08-232-463-14	Sequence 14, Appli	
9	37.4	5.0	6503	4	US-09-404-650-12	Sequence 12, Appli	
10	35.8	4.8	289	4	US-09-007-005-17	Sequence 17, Appli	
11	35.4	4.7	512	2	US-09-244-796-17	Sequence 17, Appli	
12	35.4	4.7	512	2	US-08-967-101-36	Sequence 36, Appli	
13	35.4	4.7	512	2	US-08-592-541-36	Sequence 36, Appli	
14	35.4	4.7	512	2	US-09-124-698-36	Sequence 36, Appli	
15	35.4	4.7	512	3	US-09-127-480-36	Sequence 36, Appli	
16	35.4	4.7	512	4	US-08-496-841C-36	Sequence 36, Appli	
17	35.4	4.7	512	4	US-09-124-523-36	Sequence 36, Appli	
18	35.4	4.7	512	4	US-09-007-005-32	Sequence 32, Appli	
19	34.6	4.6	248	4	US-09-244-796-32	Sequence 32, Appli	
20	34.6	4.6	248	4	US-09-007-005-3	Sequence 3, Appli	
21	34.6	4.6	277	4	US-09-244-796-3	Sequence 3, Appli	
22	34.6	4.6	277	4	US-08-471-033-39	Sequence 39, Appli	
23	34.2	4.6	1241	1	US-08-471-033-42	Sequence 42, Appli	
24	34.2	4.6	1241	2	US-08-471-044-39	Sequence 39, Appli	
25	34.2	4.6	1241	2	US-08-471-044-42	Sequence 42, Appli	
26	34.2	4.6	1241	2	US-08-471-044-42	Sequence 42, Appli	
27	34.2	4.6	1241	2	US-08-463-463A-39	Sequence 39, Appli	

Query Match 7.0%; Score 52.8; DB 4; Length 1191;
Best Local Similarity 31.4%; Pred. No. 3.8e-07;
Matches 100; Conservative 44; Mismatches 171; Indels 3; Gaps 1;

562 GTNGGNGARGCNGGNCNGAYTTNGAYWS---NYTNGTNGCNTGGGARNSMGNGNTAY 618

583 GGGGACATTATCCAGAAGGATTTCTTAGACGCTATTACAATCTGACCCCTGAAGACCATG 642

DB	619	WSNGAYTNYNYNTGSGAYTTTYYNGAYGTGCCNTTAAAYCARACNCTNNAARGAYT	678
QY	643	ATGGGCATAGAAATGGGTCCATCGCTTTTGTCCCTCAGCGCGCGTTGTGTGATGAAAAACAGAC	702
DB	679	YTNYTNYNGCTGGYINGGMMGCAYGCCACNGTMSNTTGTNTYNNMGNCNCAR	738
QY	703	TCAGACATGTTTCATCAAT	720
DB	739	GAYGAYGCNTTGTGNCAY	756

RESULT 6
US-09-459-133-1
; Sequence 1, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS

```

: PRIOR FILING DATE: 1998-12-10
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1532
: TYPE: DNA
: ORGANISM: Homo sapiens

```



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

Query Match 4.8%; Score 35.8; DB 4; Length 289;
Best Local Similarity 3.2%; Pred. No. 0.035;
Matches 6; Conservative 89; Mismatches 94; Indels 0; Gaps 0;

Qy	465	GGAGACGTGGGGGAAAGAGAGAGTGTGAAGCGAAGCAGCTGAAGACATCTTCCTCCT	524
		: :	
Db	47	RCRARURGRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRN	106
		: :	
Qy	525	GGGGACCACCACTGCACGGGAACGAAAGAGGTGGACCAGGAGCAGCACGACACGG	584
		: :	
Db	107	RNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRN	166
		: :	
Qy	585	GGACATTCCAGAAGGATTCTTAGACCTCTATTACAATCGACCTGAAGACCATGAT	644
		: :	
Db	167	RNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRN	226
		: ~:	
Qy	645	GGGCATAGA	653
		: ~:	
Db	227	RURGRCRGR	235
		: ~:	

```

RESULT 12
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak; Jack W.
; APPLICANT: Roberts; Richard W.
; APPLICANT: Liu; Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
UTS-09-244-796-17
```

Query Match 4.8%; Score 35.8; DB 4; Length 289;
Best Local Similarity 3.2%; pred. No. 0.055;
Matches 6; Conservative 89; Mismatches 94; Indels 0; Gaps 0;

QY 465 GCAGACGTGGGGGAAAGACAGGATGGTGAAGGGAAGCAGCTGAAGACATTTCTTCCTCT 524

[illegible]

RESULT 13
US-08-967-101-36

Sequence 36, Application US/09867101
Patent No. 5840340
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESS: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patcin Patent #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541

```

FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Pletcher, Edmund R.
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 248-7000
 TELEFAX: (617) 248-7100
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 512 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-967-101-36

Query Match 4.7%; Score 35.4; DB 2; Length 512;
Best Local Similarity 58.6%; Pred. No. 0.11;
Matches 51; Conservative 2; Mismatches 34; Indels 0; Gaps 0;

QY	54	TGCTGGCAGATGTTAGACCTTGTGCTTTAACTGTTTAACACACAGACGCCCGACTTCTGT	113
Db	301	TGATGTCGGANATTGTTNCTAGTGTYGCATTTAAAAAACAAGNAATATCTGCTNT	360
QY	114	ATCCAGCCGAGGCTTCAGAGTTTCCAAA	140
Db	361	ATGCTNAGAGCTNTTCAGTTTCAAA	387

RESULT 14

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-124-698-36

Query Match 4.7%; Score 35.4; DB 3;
Best Local Similarity 58.6%; Pred. No. 0.11;
Matches 51; Conservative 2; Mismatches 34;

Qy 54 TGGTGGCAGATGTAGACCTTTGTGCTTAACTGTTTAAACACACAC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TGATGTGGANATKGTNTGCTGTTGTCATATTTAAAAAACA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 114 ATGCAGCGAGGTCTTAGAGTTTCCAAA 140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 ATGCTNAGAGCTTNTCAGTTTTCARA 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: April 11, 2003, 22:48:31
Job time : 33.7377 secs

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 21:30:39 ; Search time 48.0484 Seconds
(without alignments)
13710.183 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750

Perfect score: 751
Sequence: 1 ggaccaatgccagaatctc.....gactgaactgtcttgaaga 751

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	12.1	1773	10	US-09-739-451-9
2	91	12.1	2095	9	US-09-992-598-208
3	91	12.1	2095	9	US-09-989-293A-208
4	91	12.1	2095	9	US-10-063-547-35
5	91	12.1	2095	9	US-09-989-735-208
6	91	12.1	2095	9	US-09-990-444-208
7	91	12.1	2095	9	US-09-989-730-208
8	91	12.1	2095	9	US-09-990-436-208
9	91	12.1	2095	9	US-09-991-181-208
10	91	12.1	2095	9	US-09-993-687-208
11	91	12.1	2095	9	US-09-989-734-208
12	91	12.1	2095	9	US-09-997-653-208
13	91	12.1	2095	9	US-10-174-590-161
14	91	12.1	2095	9	US-10-176-758-161
15	91	12.1	2095	9	US-10-063-616-35
16	91	12.1	2095	9	US-10-175-737-161
17	91	12.1	2095	9	US-09-993-667-208
18	91	12.1	2095	9	US-10-063-502-35
19	91	12.1	2095	9	US-10-173-706-161

20	91	12.1	2095	9	US-10-175-738-161	Sequence 161, App
21	91	12.1	2095	9	US-10-175-752-161	Sequence 161, App
22	91	12.1	2095	9	US-10-176-482-161	Sequence 161, App
23	91	12.1	2095	9	US-10-176-757-161	Sequence 161, App
24	91	12.1	2095	9	US-10-176-913-161	Sequence 161, App
25	91	12.1	2095	9	US-10-180-552-161	Sequence 161, App
26	91	12.1	2095	9	US-10-180-557-161	Sequence 161, App
27	91	12.1	2095	9	US-09-990-438-208	Sequence 208, App
28	91	12.1	2095	9	US-09-990-562-208	Sequence 208, App
29	91	12.1	2095	9	US-09-997-428-208	Sequence 208, App
30	91	12.1	2095	9	US-09-997-666-208	Sequence 208, App
31	91	12.1	2095	9	US-10-173-700-161	Sequence 161, App
32	91	12.1	2095	9	US-10-174-572-161	Sequence 161, App
33	91	12.1	2095	9	US-10-174-579-161	Sequence 161, App
34	91	12.1	2095	9	US-10-174-582-161	Sequence 161, App
35	91	12.1	2095	9	US-10-174-588-161	Sequence 161, App
36	91	12.1	2095	9	US-10-175-733-161	Sequence 161, App
37	91	12.1	2095	9	US-10-175-740-161	Sequence 161, App
38	91	12.1	2095	9	US-10-175-743-161	Sequence 161, App
39	91	12.1	2095	9	US-10-176-488-161	Sequence 161, App
40	91	12.1	2095	9	US-10-176-492-161	Sequence 161, App
41	91	12.1	2095	9	US-10-176-747-161	Sequence 161, App
42	91	12.1	2095	9	US-10-176-750-161	Sequence 161, App
43	91	12.1	2095	9	US-10-176-985-161	Sequence 161, App
44	91	12.1	2095	9	US-10-176-987-161	Sequence 161, App
45	91	12.1	2095	9	US-10-176-991-161	Sequence 161, App

ALIGNMENTS

RESULT 1

US-09-739-451-9
; Sequence 9, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-739-451-9

Query Match	12.1%	Score 91;	DB 10;	Length 1773;
Best Local Similarity	53.3%	Pred. No. 4.7e+20;		
Mismatches	221;	Conservative 0;	Mismatches 185;	Indels 9; Gaps 1;
QY	344	TTTACAGAAGACGGGAACCTTCCTTAAGCTCCAGATACAGACTGCAGGACACACCTC	403	
Db	157	TTTACAGAAGACTTTCACACTTCGAGAGCATTCACAACTGCTCTCATCAAAATC	216	
QY	404	CTTCTCTGCTGCTGGTGACCTCATCCACAAACAGTTGGTGTAGGCCATGGCCATCC	463	
Db	217	CATTCTGTGTCATCTGGTGACCTCCACCTTCAGATGTAAAGCCAGGCGGCATTA	276	
QY	464	GGCAGAGCTGGGGAAGAGAGGATGGTGAAGGAAGCAGCTGAAGACATCTTCCTCC	523	
Db	277	GAGTTACTTGGGTGAAAAAAGTCTTGTGGGATATGAGGTCTTACATTTTCTTAT	336	
QY	524	TGGGACACACACAGCTGCAGGGAACCAAGAGG-----TGGACACGAGAGACC	574	

DB 337 TAGGCAAGAGCTGAAAGAGGAGACAAATATGTTGGCATTGTCTTAGAGGATGAACACC 396
QY 575 AGCGACACGGGACATTATCCAGAGGATTTCTTAGACGCTCTATTACAATCTGACCTGA 634
DB 397 TTCATTATGGTGACATATCCGACAGATTTTATAGACACATATATAACCTGACCTGA 456
QY 635 AGACCATGATGGGATAGATGGGTCCATCGCTTTTCTCAGGGCGGCTTTGTGTATGA 694
DB 457 AAACCATTTGGCATTCAGGTGGTAACTGAGTTTCCGCCCAATGCCAAGTACGTAATGA 516
QY 695 AAACAGACATCAGACATGTTCAATCAATGTTGACTACTGACTGAACCTCTCTGAA 749
DB 517 AGACAGACACTGATGTTTTCATCACTACTGCGCAATTTAGTGAAGTATCTTTTAAA 571

RESULT 2
US-09-992-598-208
Sequence 208 Application US/09992598
Patent No. US20020160384A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secured and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C20
CURRENT APPLICATION NUMBER: US/09/992,598
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908

1 PRIOR FILING DATE: 1998-06-18
2 PRIOR APPLICATION NUMBER: 60/089947
3 PRIOR FILING DATE: 1998-06-19
4 PRIOR APPLICATION NUMBER: 60/089948
5 PRIOR FILING DATE: 1998-06-19
6 PRIOR APPLICATION NUMBER: 60/089952
7 PRIOR FILING DATE: 1998-06-19
8 PRIOR APPLICATION NUMBER: 60/090246
9 PRIOR FILING DATE: 1998-06-22
10 PRIOR APPLICATION NUMBER: 60/090252
11 PRIOR FILING DATE: 1998-06-22
12 PRIOR APPLICATION NUMBER: 60/090254
13 PRIOR FILING DATE: 1998-06-22
14 PRIOR APPLICATION NUMBER: 60/090349
15 PRIOR FILING DATE: 1998-06-23
16 PRIOR APPLICATION NUMBER: 60/090355
17 PRIOR FILING DATE: 1998-06-23
18 PRIOR APPLICATION NUMBER: 60/090429
19 PRIOR FILING DATE: 1998-06-24
20 PRIOR APPLICATION NUMBER: 60/090431
21 PRIOR FILING DATE: 1998-06-24
22 PRIOR APPLICATION NUMBER: 60/090435
23 PRIOR FILING DATE: 1998-06-24
24 PRIOR APPLICATION NUMBER: 60/090444
25 PRIOR FILING DATE: 1998-06-24
26 PRIOR APPLICATION NUMBER: 60/090445
27 PRIOR FILING DATE: 1998-06-24
28 PRIOR APPLICATION NUMBER: 60/090472
29 PRIOR FILING DATE: 1998-06-24
30 PRIOR APPLICATION NUMBER: 60/090535
31 PRIOR FILING DATE: 1998-06-24
32 PRIOR APPLICATION NUMBER: 60/090540
33 PRIOR FILING DATE: 1998-06-24
34 PRIOR APPLICATION NUMBER: 60/090542
35 PRIOR FILING DATE: 1998-06-24
36 PRIOR APPLICATION NUMBER: 60/090557
37 PRIOR FILING DATE: 1998-06-24
38 PRIOR APPLICATION NUMBER: 60/090676
39 PRIOR FILING DATE: 1998-06-25
40 PRIOR APPLICATION NUMBER: 60/090678
41 PRIOR FILING DATE: 1998-06-25
42 PRIOR APPLICATION NUMBER: 60/090690
43 PRIOR FILING DATE: 1998-06-25
44 PRIOR APPLICATION NUMBER: 60/090694
45 PRIOR FILING DATE: 1998-06-25
46 PRIOR APPLICATION NUMBER: 60/090695
47 PRIOR FILING DATE: 1998-06-25
48 PRIOR APPLICATION NUMBER: 60/090696
49 PRIOR FILING DATE: 1998-06-25
50 PRIOR APPLICATION NUMBER: 60/090862
51 PRIOR FILING DATE: 1998-06-26
52 PRIOR APPLICATION NUMBER: 60/090863
53 PRIOR FILING DATE: 1998-06-26
54 PRIOR APPLICATION NUMBER: 60/091360
55 PRIOR FILING DATE: 1998-07-01
56 PRIOR APPLICATION NUMBER: 60/091478
57 PRIOR FILING DATE: 1998-07-02
58 PRIOR APPLICATION NUMBER: 60/091544
59 PRIOR FILING DATE: 1998-07-01
60 PRIOR APPLICATION NUMBER: 60/091519
61 PRIOR FILING DATE: 1998-07-02
62 PRIOR APPLICATION NUMBER: 60/091626
63 PRIOR FILING DATE: 1998-07-02
64 PRIOR APPLICATION NUMBER: 60/091633
65 PRIOR FILING DATE: 1998-07-02
66 PRIOR APPLICATION NUMBER: 60/091978
67 PRIOR FILING DATE: 1998-07-07
68 PRIOR APPLICATION NUMBER: 60/091982
69 PRIOR FILING DATE: 1998-07-07
70 PRIOR APPLICATION NUMBER: 60/092182
71 PRIOR FILING DATE: 1998-07-09

Query Match

12.1%; Score 91; DB 9; Length 2095;

Best Local Similarity 53.3%; Pred. No. 5,3e-20;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;
QY 344 TTTACAAGAAAGACGGGAACACTTCCTTAAGCTCCAGATACAGACTGCAGGAGACACCTC 403
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 TTTACAGACAAGACITTCACITTCAGACTTCGAGAGACATTCAAACTGCTCTCATCAAAATC 553
QY 404 CTTCTCTCGTCTGCTGGTGACCTCCACACAAACAGTTGGCTGAGCGCATGGCCATCC 463
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 CATTTCCTGGTCAATTCUGGTGACCTCCACCCCTTCAGATGTGAAGCCAGGAGGCCATTA 613
QY 464 GCGAGACGTGGGGAAGAGAGAGATGGTGAAGGGAAGACAGCTGAAGACATTCCTCTCC 523
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 GAGTTACTTGGGTGAAGAAAAAGTCTTGGTGGGATATGAGGTTCTTACATTTTCTTAT 673
QY 524 TGGGGACACACAGCAGCTGCGGGGAAACGAAAGAGG-----TGGACAGAGAGGCC 574
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 674 TAGGCCAAGAGGCTGAAAGAGGAGACAAATGTTGGCATTGTCTCTTAGAGATGAACACC 733
QY 575 AGCGACACGGGACATTCACAGAGGATTTCTAGAGGTCTATTACAACTGACCCCTGA 634
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 734 TTTCTTATGTCACATAATCCGACAGATTTTATAGACACATATATATACCTGACCTTGA 793
QY 635 AGACCATGATGGGCATAGAAATGGGTCCATCGCTTTTTCCTCAGCGCGCTTTGTGATGA 694
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 AAACCATTTATGCCATTCAGGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCAATGTTGACTATCTCAGTGAACCTGCTTCTGAA 749
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAA 908

RESULT 3

US-09-989-293A-208
; Sequence 208, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311

PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478

1 PRIOR FILING DATE: 1997-10-17
2 PRIOR APPLICATION NUMBER: 60/065186
3 PRIOR FILING DATE: 1997-11-12
4 PRIOR APPLICATION NUMBER: 60/065311
5 PRIOR FILING DATE: 1997-11-13
6 PRIOR APPLICATION NUMBER: 60/066770
7 PRIOR FILING DATE: 1997-11-24
8 PRIOR APPLICATION NUMBER: 60/075945
9 PRIOR FILING DATE: 1998-02-25
10 PRIOR APPLICATION NUMBER: 60/078910
11 PRIOR FILING DATE: 1998-03-20
12 PRIOR APPLICATION NUMBER: 60/083322
13 PRIOR FILING DATE: 1998-04-28
14 PRIOR APPLICATION NUMBER: 60/084600
15 PRIOR FILING DATE: 1998-05-07
16 PRIOR APPLICATION NUMBER: 60/087106
17 PRIOR FILING DATE: 1998-05-28
18 PRIOR APPLICATION NUMBER: 60/087607
19 PRIOR FILING DATE: 1998-06-02
20 PRIOR APPLICATION NUMBER: 60/087609
21 PRIOR FILING DATE: 1998-06-02
22 PRIOR APPLICATION NUMBER: 60/087759
23 PRIOR FILING DATE: 1998-06-02
24 PRIOR APPLICATION NUMBER: 60/087827
25 PRIOR FILING DATE: 1998-06-03
26 PRIOR APPLICATION NUMBER: 60/088021
27 PRIOR FILING DATE: 1998-06-04
28 PRIOR APPLICATION NUMBER: 60/088025
29 PRIOR FILING DATE: 1998-06-04
30 PRIOR APPLICATION NUMBER: 60/088026
31 PRIOR FILING DATE: 1998-06-04
32 PRIOR APPLICATION NUMBER: 60/088028
33 PRIOR FILING DATE: 1998-06-04
34 PRIOR APPLICATION NUMBER: 60/088029
35 PRIOR FILING DATE: 1998-06-04
36 PRIOR APPLICATION NUMBER: 60/088030
37 PRIOR FILING DATE: 1998-06-04
38 PRIOR APPLICATION NUMBER: 60/088033
39 PRIOR FILING DATE: 1998-06-04
40 PRIOR APPLICATION NUMBER: 60/088326
41 PRIOR FILING DATE: 1998-06-04
42 PRIOR APPLICATION NUMBER: 60/088167
43 PRIOR FILING DATE: 1998-06-05
44 PRIOR APPLICATION NUMBER: 60/088202
45 PRIOR FILING DATE: 1998-06-05
46 PRIOR APPLICATION NUMBER: 60/088212
47 PRIOR FILING DATE: 1998-06-05
48 PRIOR APPLICATION NUMBER: 60/088217
49 PRIOR FILING DATE: 1998-06-05
50 PRIOR APPLICATION NUMBER: 60/088655
51 PRIOR FILING DATE: 1998-06-09
52 PRIOR APPLICATION NUMBER: 60/088734
53 PRIOR FILING DATE: 1998-06-10
54 PRIOR APPLICATION NUMBER: 60/088738
55 PRIOR FILING DATE: 1998-06-10
56 PRIOR APPLICATION NUMBER: 60/088742
57 PRIOR FILING DATE: 1998-06-10
58 PRIOR APPLICATION NUMBER: 60/088810
59 PRIOR FILING DATE: 1998-06-10
60 PRIOR APPLICATION NUMBER: 60/088824
61 PRIOR FILING DATE: 1998-06-10
62 PRIOR APPLICATION NUMBER: 60/088826
63 PRIOR FILING DATE: 1998-06-10
64 PRIOR APPLICATION NUMBER: 60/088858
65 PRIOR FILING DATE: 1998-06-11
66 PRIOR APPLICATION NUMBER: 60/088861
67 PRIOR FILING DATE: 1998-06-11
68 PRIOR APPLICATION NUMBER: 60/088876
69 PRIOR FILING DATE: 1998-06-11
70 PRIOR APPLICATION NUMBER: 60/089105
71 PRIOR FILING DATE: 1998-06-12
72 PRIOR APPLICATION NUMBER: 60/089440
73 PRIOR FILING DATE: 1998-06-16
74 PRIOR APPLICATION NUMBER: 60/089512
75 PRIOR FILING DATE: 1998-06-16
76 PRIOR APPLICATION NUMBER: 60/089514
77 PRIOR FILING DATE: 1998-06-16
78 PRIOR APPLICATION NUMBER: 60/089532
79 PRIOR FILING DATE: 1998-06-17
80 PRIOR APPLICATION NUMBER: 60/089538
81 PRIOR FILING DATE: 1998-06-17
82 PRIOR APPLICATION NUMBER: 60/089598
83 PRIOR FILING DATE: 1998-06-17
84 PRIOR APPLICATION NUMBER: 60/089599
85 PRIOR FILING DATE: 1998-06-17
86 PRIOR APPLICATION NUMBER: 60/089600
87 PRIOR FILING DATE: 1998-06-17
88 PRIOR APPLICATION NUMBER: 60/089653
89 PRIOR FILING DATE: 1998-06-17
90 PRIOR APPLICATION NUMBER: 60/089801
91 PRIOR FILING DATE: 1998-06-18
92 PRIOR APPLICATION NUMBER: 60/089907
93 PRIOR FILING DATE: 1998-06-18
94 PRIOR APPLICATION NUMBER: 60/089908
95 PRIOR FILING DATE: 1998-06-18
96 PRIOR APPLICATION NUMBER: 60/089947
97 PRIOR FILING DATE: 1998-06-19
98 PRIOR APPLICATION NUMBER: 60/089948
99 PRIOR FILING DATE: 1998-06-19
100 PRIOR APPLICATION NUMBER: 60/089952
101 PRIOR FILING DATE: 1998-06-19
102 PRIOR APPLICATION NUMBER: 60/090246
103 PRIOR FILING DATE: 1998-06-22
104 PRIOR APPLICATION NUMBER: 60/090252
105 PRIOR FILING DATE: 1998-06-22
106 PRIOR APPLICATION NUMBER: 60/090254
107 PRIOR FILING DATE: 1998-06-22
108 PRIOR APPLICATION NUMBER: 60/090349
109 PRIOR FILING DATE: 1998-06-23
110 PRIOR APPLICATION NUMBER: 60/090355
111 PRIOR FILING DATE: 1998-06-23
112 PRIOR APPLICATION NUMBER: 60/090429
113 PRIOR FILING DATE: 1998-06-24
114 PRIOR APPLICATION NUMBER: 60/090431
115 PRIOR FILING DATE: 1998-06-24
116 PRIOR APPLICATION NUMBER: 60/090435
117 PRIOR FILING DATE: 1998-06-24
118 PRIOR APPLICATION NUMBER: 60/090444
119 PRIOR FILING DATE: 1998-06-24
120 PRIOR APPLICATION NUMBER: 60/090445
121 PRIOR FILING DATE: 1998-06-24
122 PRIOR APPLICATION NUMBER: 60/090472
123 PRIOR FILING DATE: 1998-06-24
124 PRIOR APPLICATION NUMBER: 60/090535
125 PRIOR FILING DATE: 1998-06-24
126 PRIOR APPLICATION NUMBER: 60/090540
127 PRIOR FILING DATE: 1998-06-24
128 PRIOR APPLICATION NUMBER: 60/090542
129 PRIOR FILING DATE: 1998-06-24
130 PRIOR APPLICATION NUMBER: 60/090557
131 PRIOR FILING DATE: 1998-06-24
132 PRIOR APPLICATION NUMBER: 60/090676
133 PRIOR FILING DATE: 1998-06-25
134 PRIOR APPLICATION NUMBER: 60/090678
135 PRIOR FILING DATE: 1998-06-25
136 PRIOR APPLICATION NUMBER: 60/090690
137 PRIOR FILING DATE: 1998-06-25
138 PRIOR APPLICATION NUMBER: 60/090694
139 PRIOR FILING DATE: 1998-06-25
140 PRIOR APPLICATION NUMBER: 60/090695
141 PRIOR FILING DATE: 1998-06-25
142 PRIOR APPLICATION NUMBER: 60/090696
143 PRIOR FILING DATE: 1998-06-25
144 PRIOR APPLICATION NUMBER: 60/090862
145 PRIOR FILING DATE: 1998-06-26
146 PRIOR APPLICATION NUMBER: 60/090863

Publication No. US20020197674A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P27301C69
CURRENT APPLICATION NUMBER: US/09/989,730
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088555
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429

```
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090431
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090435
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090444
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090445
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090472
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090535
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090540
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090542
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090557
? PRIOR FILING DATE: 1998-06-24
? PRIOR APPLICATION NUMBER: 60/090676
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090678
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090690
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090694
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090695
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090696
? PRIOR FILING DATE: 1998-06-25
? PRIOR APPLICATION NUMBER: 60/090862
? PRIOR FILING DATE: 1998-06-26
? PRIOR APPLICATION NUMBER: 60/090863
? PRIOR FILING DATE: 1998-06-26
? PRIOR APPLICATION NUMBER: 60/091360
? PRIOR FILING DATE: 1998-07-01
? PRIOR APPLICATION NUMBER: 60/091478
? PRIOR FILING DATE: 1998-07-02
? PRIOR APPLICATION NUMBER: 60/091544
? PRIOR FILING DATE: 1998-07-01
? PRIOR APPLICATION NUMBER: 60/091519
? PRIOR FILING DATE: 1998-07-02
? PRIOR APPLICATION NUMBER: 60/091626
? PRIOR FILING DATE: 1998-07-02
? PRIOR APPLICATION NUMBER: 60/091633
? PRIOR FILING DATE: 1998-07-02
? PRIOR APPLICATION NUMBER: 60/091978
? PRIOR FILING DATE: 1998-07-07
? PRIOR APPLICATION NUMBER: 60/091982
? PRIOR FILING DATE: 1998-07-07
? PRIOR APPLICATION NUMBER: 60/092182
? PRIOR FILING DATE: 1998-07-09

Query Match 12.1%; Score 91; DB 9; Length 2095;
Best Local Similarity 53.3%; Pred. No. 5.3e-20;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;

QY 344 TTTACAAAGAGCGGGAAGTCTTCTTAAGCTCCAGATACAGACTGCAGGACAGACACCTC 403
DB 494 TTTACAGACAAGACTTTTCACTTCCAGACTTCGAGAGCAATCAACTGCTCTCATCAAAATC 553
QY 404 CCTTCTCGTCCCTGGTGGTCACTTCCCAACAAAGTGGCTGAGCGCATGGCCATCC 463
DB 554 CATTTCTGGTCACTTCTGGTGGTCACTTCCCAACAAAGTGGCTGAGCGCATGGCCATTA 613
QY 464 GCGAGAGTGGGGAAGAGAGAGGATGGTGAAGGAAGAGAGCTGAAGACATCTTCTCTCC 523
DB 614 GAGTTACTTGGGGTGAAGAAAGTCTTGGTGGGATATGAGGTTCTTACATTTTCTTAT 673
QY 524 TGGGGACCAACAGCAGTGCAGCGGAAACGAAAGG-----TGGACCAAGAGAGCC 574
DB 674 TAGGCCAAGAGGCTGAAGAGGAAGACAAATGTTGGCATTGCTTCTTAGAGGATGAACACC 733

QY 575 AGCGACAGCGGACACATTATCCAGAGAGGATTTCTAGAGCTCTATTACAATCTGACCTGA 634
DB 734 TTTCTTTATGGTGACATAATCCGACAAAGATTTTGTAGACACATATAATAAAGCTGACCTGA 793
QY 635 AGACCATGATGGGCATAGAAATGGGTCCCATCGCTTTTGTCTTCAGGCGGCTTTGTGATGA 694
DB 794 AAACCATATTGSCATTGAGTGGGTAACTGAGTTTGTCCCAATGCCAAGTACCTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAA 749
DB 854 AGACAGACACTGATGTTTTCATCAATACTACTGCGCAATTTAGTAGAGTATCTTTTAAA 908

RESULT 8
US-09-990-436-208
; Sequence 208, Application US/09990436
; Publication No. US20020198148A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C14
; CURRENT APPLICATION NUMBER: US/09/990,436
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
```

1	PRIOR FILING DATE: 1998-06-02	1	PRIOR APPLICATION NUMBER: 60/089944	1
2	PRIOR APPLICATION NUMBER: 60/087827	2	PRIOR FILING DATE: 1998-06-19	2
3	PRIOR FILING DATE: 1998-06-03	3	PRIOR APPLICATION NUMBER: 60/089948	3
4	PRIOR APPLICATION NUMBER: 60/088021	4	PRIOR FILING DATE: 1998-06-19	4
5	PRIOR FILING DATE: 1998-06-04	5	PRIOR APPLICATION NUMBER: 60/089952	5
6	PRIOR APPLICATION NUMBER: 60/088025	6	PRIOR FILING DATE: 1998-06-19	6
7	PRIOR FILING DATE: 1998-06-04	7	PRIOR APPLICATION NUMBER: 60/090246	7
8	PRIOR APPLICATION NUMBER: 60/088026	8	PRIOR FILING DATE: 1998-06-22	8
9	PRIOR FILING DATE: 1998-06-04	9	PRIOR APPLICATION NUMBER: 60/090252	9
10	PRIOR APPLICATION NUMBER: 60/088028	10	PRIOR FILING DATE: 1998-06-22	10
11	PRIOR FILING DATE: 1998-06-04	11	PRIOR APPLICATION NUMBER: 60/090254	11
12	PRIOR APPLICATION NUMBER: 60/088029	12	PRIOR FILING DATE: 1998-06-22	12
13	PRIOR FILING DATE: 1998-06-04	13	PRIOR APPLICATION NUMBER: 60/090349	13
14	PRIOR APPLICATION NUMBER: 60/088030	14	PRIOR FILING DATE: 1998-06-22	14
15	PRIOR FILING DATE: 1998-06-04	15	PRIOR APPLICATION NUMBER: 60/090355	15
16	PRIOR APPLICATION NUMBER: 60/088033	16	PRIOR FILING DATE: 1998-06-23	16
17	PRIOR FILING DATE: 1998-06-04	17	PRIOR APPLICATION NUMBER: 60/090429	17
18	PRIOR APPLICATION NUMBER: 60/088326	18	PRIOR FILING DATE: 1998-06-24	18
19	PRIOR FILING DATE: 1998-06-04	19	PRIOR APPLICATION NUMBER: 60/090431	19
20	PRIOR APPLICATION NUMBER: 60/088167	20	PRIOR FILING DATE: 1998-06-24	20
21	PRIOR FILING DATE: 1998-06-05	21	PRIOR APPLICATION NUMBER: 60/090435	21
22	PRIOR APPLICATION NUMBER: 60/088202	22	PRIOR FILING DATE: 1998-06-24	22
23	PRIOR FILING DATE: 1998-06-05	23	PRIOR APPLICATION NUMBER: 60/090444	23
24	PRIOR APPLICATION NUMBER: 60/088212	24	PRIOR FILING DATE: 1998-06-24	24
25	PRIOR FILING DATE: 1998-06-05	25	PRIOR APPLICATION NUMBER: 60/090445	25
26	PRIOR APPLICATION NUMBER: 60/088217	26	PRIOR FILING DATE: 1998-06-24	26
27	PRIOR FILING DATE: 1998-06-05	27	PRIOR APPLICATION NUMBER: 60/090472	27
28	PRIOR APPLICATION NUMBER: 60/088555	28	PRIOR FILING DATE: 1998-06-24	28
29	PRIOR FILING DATE: 1998-06-09	29	PRIOR APPLICATION NUMBER: 60/090535	29
30	PRIOR APPLICATION NUMBER: 60/088734	30	PRIOR FILING DATE: 1998-06-24	30
31	PRIOR FILING DATE: 1998-06-10	31	PRIOR APPLICATION NUMBER: 60/090540	31
32	PRIOR APPLICATION NUMBER: 60/088738	32	PRIOR FILING DATE: 1998-06-24	32
33	PRIOR FILING DATE: 1998-06-10	33	PRIOR APPLICATION NUMBER: 60/090542	33
34	PRIOR APPLICATION NUMBER: 60/088742	34	PRIOR FILING DATE: 1998-06-24	34
35	PRIOR FILING DATE: 1998-06-10	35	PRIOR APPLICATION NUMBER: 60/090557	35
36	PRIOR APPLICATION NUMBER: 60/088810	36	PRIOR FILING DATE: 1998-06-24	36
37	PRIOR FILING DATE: 1998-06-10	37	PRIOR APPLICATION NUMBER: 60/090576	37
38	PRIOR APPLICATION NUMBER: 60/088824	38	PRIOR FILING DATE: 1998-06-25	38
39	PRIOR FILING DATE: 1998-06-10	39	PRIOR APPLICATION NUMBER: 60/090678	39
40	PRIOR APPLICATION NUMBER: 60/088826	40	PRIOR FILING DATE: 1998-06-25	40
41	PRIOR FILING DATE: 1998-06-10	41	PRIOR APPLICATION NUMBER: 60/090690	41
42	PRIOR APPLICATION NUMBER: 60/088858	42	PRIOR FILING DATE: 1998-06-25	42
43	PRIOR FILING DATE: 1998-06-11	43	PRIOR APPLICATION NUMBER: 60/090694	43
44	PRIOR APPLICATION NUMBER: 60/088861	44	PRIOR FILING DATE: 1998-06-25	44
45	PRIOR FILING DATE: 1998-06-11	45	PRIOR APPLICATION NUMBER: 60/090695	45
46	PRIOR APPLICATION NUMBER: 60/088876	46	PRIOR FILING DATE: 1998-06-25	46
47	PRIOR FILING DATE: 1998-06-11	47	PRIOR APPLICATION NUMBER: 60/090862	47
48	PRIOR APPLICATION NUMBER: 60/089105	48	PRIOR FILING DATE: 1998-06-26	48
49	PRIOR FILING DATE: 1998-06-12	49	PRIOR APPLICATION NUMBER: 60/090863	49
50	PRIOR APPLICATION NUMBER: 60/089440	50	PRIOR FILING DATE: 1998-06-26	50
51	PRIOR FILING DATE: 1998-06-16	51	PRIOR APPLICATION NUMBER: 60/091360	51
52	PRIOR APPLICATION NUMBER: 60/089512	52	PRIOR FILING DATE: 1998-07-01	52
53	PRIOR FILING DATE: 1998-06-16	53	PRIOR APPLICATION NUMBER: 60/091478	53
54	PRIOR APPLICATION NUMBER: 60/089514	54	PRIOR FILING DATE: 1998-07-02	54
55	PRIOR FILING DATE: 1998-06-16	55	PRIOR APPLICATION NUMBER: 60/091544	55
56	PRIOR APPLICATION NUMBER: 60/089532	56	PRIOR FILING DATE: 1998-07-01	56
57	PRIOR FILING DATE: 1998-06-17	57	PRIOR APPLICATION NUMBER: 60/091519	57
58	PRIOR APPLICATION NUMBER: 60/089538	58	PRIOR FILING DATE: 1998-07-02	58
59	PRIOR FILING DATE: 1998-06-17	59	PRIOR APPLICATION NUMBER: 60/091626	59
60	PRIOR APPLICATION NUMBER: 60/089598	60	PRIOR FILING DATE: 1998-07-02	60
61	PRIOR FILING DATE: 1998-06-17	61		

Query Match 12.18; Score 91; DB 9; Length 2095;
Best Local Similarity 53.38; Pred. No. 5.3e-20;

RESULT 9

APPLICANT: Stewart, Timothy

APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730P1C11
CURRENT APPLICATION NUMBER: US/09/993,687
CURRENT FILING DATE: 2002-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678

1	PRIOR APPLICATION NUMBER: 60/090444	
2	PRIOR FILING DATE: 1998-06-24	
3	PRIOR APPLICATION NUMBER: 60/090445	
4	PRIOR FILING DATE: 1998-06-24	
5	PRIOR APPLICATION NUMBER: 60/090472	
6	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090535	
8	PRIOR FILING DATE: 1998-06-24	
9	PRIOR APPLICATION NUMBER: 60/090540	
10	PRIOR FILING DATE: 1998-06-24	
11	PRIOR APPLICATION NUMBER: 60/090542	
12	PRIOR FILING DATE: 1998-06-24	
13	PRIOR APPLICATION NUMBER: 60/090557	
14	PRIOR FILING DATE: 1998-06-24	
15	PRIOR APPLICATION NUMBER: 60/090676	
16	PRIOR FILING DATE: 1998-06-25	
17	PRIOR APPLICATION NUMBER: 60/090678	
18	PRIOR FILING DATE: 1998-06-25	
19	PRIOR APPLICATION NUMBER: 60/090690	
20	PRIOR FILING DATE: 1998-06-25	
21	PRIOR APPLICATION NUMBER: 60/090694	
22	PRIOR FILING DATE: 1998-06-25	
23	PRIOR APPLICATION NUMBER: 60/090695	
24	PRIOR FILING DATE: 1998-06-25	
25	PRIOR APPLICATION NUMBER: 60/090696	
26	PRIOR FILING DATE: 1998-06-25	
27	PRIOR APPLICATION NUMBER: 60/091360	
28	PRIOR FILING DATE: 1998-07-01	
29	PRIOR APPLICATION NUMBER: 60/091478	
30	PRIOR FILING DATE: 1998-07-02	
31	PRIOR APPLICATION NUMBER: 60/091544	
32	PRIOR FILING DATE: 1998-07-01	
33	PRIOR APPLICATION NUMBER: 60/091519	
34	PRIOR FILING DATE: 1998-07-02	
35	PRIOR APPLICATION NUMBER: 60/091626	
36	PRIOR FILING DATE: 1998-07-02	
37	PRIOR APPLICATION NUMBER: 60/091633	
38	PRIOR FILING DATE: 1998-07-02	
39	PRIOR APPLICATION NUMBER: 60/091978	
40	PRIOR FILING DATE: 1998-07-07	
41	PRIOR APPLICATION NUMBER: 60/091982	
42	PRIOR FILING DATE: 1998-07-07	
43	PRIOR APPLICATION NUMBER: 60/092182	
44	PRIOR FILING DATE: 1998-07-09	

Query Match 12.1%; Score 91; DB 9; Length 2095;
Best Local Similarity 53.3%; Pred. No. 5.3e-20;

Qy	344	TTTACAGAAAGACGGGAACCTCTTAAGCTCCAGATACAGACTGCAGGCAGACACCTC	403
Db	494	TTTTACACAAAGACTTTTCACCTTCACACTTCGAGAGCATTCAAAGTGTCTCATCAAAATC	553
Qy	404	CTTCTCTCGTCTGCTGGTGTACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCC	463
Db	554	CATTTCTGGTCAATTCCTGGTGAACCTCCACCCCTTCAGATGTGAAGCCAGCGAGGCCATT	613
Qy	464	GGCAGACGTGGGGAAACAGAGGATGGTGAAGGAAAGCAGCTCGAAGACATTTCTCCTCC	523
Db	614	GAGTTACTTGGGTGCAAAAAGAGTCTTGGTGGGGATATGAGTTCTTACATTTTCTTAT	673
Qy	524	TGGGGACACCAGCAGTGCAGCGGAAACGAAAGAGG-----TGGACCGAGGAGCC	574
Db	674	TAGCCCAAGAGGCTGAAAAGGAAGACAAATGTTGGCATTTCTCTTAGAGGATGAACCC	733
Qy	575	AGCCACACGGGGACATTATCCAGAGGATTTCTCTAGACGTCATTACAATCTGACCCCTGA	634
Db	734	TTCTTTATGGTGAACATAATCCGACAGATTTTTTAGACACATATAATACCTGACCTTGA	793

QY 635 AGACATGATGGCATAGATGGGTCACCGCTTTTGTCTCAGCGGCGTTCTGTGA 694
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 AAACCATATGCGCATAGGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 695 AAACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAACCTGCTTCTGAA 749
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 AGACAGACACTGATGTTTTCATCACTACTGCGCAATTAGTGAATATCTTTTAA 908
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12

US-09-997-653-208
; Sequence 208, Application US/09997653
; Publication No. US20030008297A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Geiber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730P1C38
; CURRENT APPLICATION NUMBER: US/09/997,653
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952

```

Query Match      12.18; Score 91; DB 9; Length 2095;
Best Local Similarity 53.3%; Pred. No. 5.3e-20;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;

QY 344 TTTTCAAGAAAGACGGGAACCTTCCTTAAGTCCCGCATAGACACTCGAGCGACACACCTC 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 TTTTACAGACAGACTTTCACCTTCACACTTCGAGAGCAATCGAACTGCTTCATCAAAATC 553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Qy	403
Db	553

```

344 TTTACAGAAAGACGGGAATCTCCTTAAGCTCCAGATACAGACTGCAGGCGAGACACCTC
      |||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||
494 TTTACAGACAAGACTTTCACCTTCAGAGAGCATCAAACTGCTCTCATCAAAAATC
    
```

```
Db 674 TAGGCCAAGAGGCTGAAAAGGAGACAAAATGTTGGCATTGTCTTAGAGGATGAACACC 733
QY 575 AGCGACACGGGACGACATATCCAGAGGATTTCTAGAGGCTCTATTACAACTCGACCCCTGA 634
Db 734 TTTCTTATGGTGACATATCCGACAGATTTTATAGACACATATAATAACCTGACCTTGA 793
QY 635 AGACCATGATGGCATAGATGGTGCCATCGCTTTTGTCTCAGGGGGGCTTTGTGATGA 694
Db 794 AAACCATTTAGCCATTGAGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCAATCTTGACTATCTGACTGAACCTGCTCTGAA 749
Db 854 AGACAGACACTGATGTTTTCATCACTACTGGCAATTTAGTGAAGTATCTTTTAAA 908

RESULT 14
US-10-176-758-161
; Sequence 161, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 161
; LENGTH: 2095
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-758-161

Query Match 12.1%; Score 91; DB 9; Length 2095;
Best Local Similarity 53.3%; Pred. No. 5.3e-20;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;

QY 344 TTACAGAGAACGCGGAACTTCCTTAAGCTCCAGATACAGACTGCAGGACAGACACCTC 403
Db 494 TTACAGACAGACTTTTCACTTCACACTTCGAGAGCAITTCAACTGCTCTCATCAAAATC 553
QY 404 CCTTCTCGTCTGCTGAGCTATCCCAACAAAGAGGATGGTGAAGGAAAGCAGCTGAACACATCTTCTCC 523
Db 614 GAGTTACTTGGGTGAAAAAAGTCTTGGTGGGATATGAGTTCTTACATTTTCTTAT 673
QY 524 TGGGACCAACGAGCTGAAAGGAAAGCAAAATGTTGGCATTGCTTAGAGGATGAACACC 733
Db 734 TTTCTTATGGTGACATATCCGACAGATTTTATAGACACATATAATAACCTGACCTTGA 793
QY 635 AGACCATGATGGCATAGATGGTGCCATCGCTTTTGTCTCAGGGGGGCTTTGTGATGA 694
Db 794 AAACCATTTAGCCATTGAGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCACTACTGGCAATTTAGTGAAGTATCTTTTAAA 908

Search completed: April 12, 2003, 04:34:49
Job time : 54.0484 secs
```

```
Db 674 TAGGCCAAGAGGCTGAAAAGGAGACAAAATGTTGGCATTGTCTTAGAGGATGAACACC 733
QY 575 AGCGACACGGGACGACATATCCAGAGGATTTCTAGAGGCTCTATTACAACTCGACCCCTGA 634
Db 734 TTTCTTATGGTGACATATCCGACAGATTTTATAGACACATATAATAACCTGACCTTGA 793
QY 635 AGACCATGATGGCATAGATGGTGCCATCGCTTTTGTCTCAGGGGGGCTTTGTGATGA 694
Db 794 AAACCATTTAGCCATTGAGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCAATCTTGACTATCTGACTGAACCTGCTCTGAA 749
Db 854 AGACAGACACTGATGTTTTCATCACTACTGGCAATTTAGTGAAGTATCTTTTAAA 908

RESULT 15
US-10-063-616-35
; Sequence 35, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 35
; LENGTH: 2095
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-616-35

Query Match 12.1%; Score 91; DB 9; Length 2095;
Best Local Similarity 53.3%; Pred. No. 5.3e-20;
Matches 221; Conservative 0; Mismatches 185; Indels 9; Gaps 1;

QY 344 TTACAGAGAACGCGGAACTTCCTTAAGCTCCAGATACAGACTGCAGGACAGACACCTC 403
Db 494 TTACAGACAGACTTTTCACTTCACACTTCGAGAGCAITTCAACTGCTCTCATCAAAATC 553
QY 404 CCTTCTCGTCTGCTGAGCTATCCCAACAAAGAGGATGGTGAAGGAAAGCAGCTGAACACATCTTCTCC 523
Db 614 GAGTTACTTGGGTGAAAAAAGTCTTGGTGGGATATGAGTTCTTACATTTTCTTAT 673
QY 524 TGGGACCAACGAGCTGAAAGGAAAGCAAAATGTTGGCATTGCTTAGAGGATGAACACC 733
Db 734 TTTCTTATGGTGACATATCCGACAGATTTTATAGACACATATAATAACCTGACCTTGA 793
QY 635 AGACCATGATGGCATAGATGGTGCCATCGCTTTTGTCTCAGGGGGGCTTTGTGATGA 694
Db 794 AAACCATTTAGCCATTGAGTGGTAACTGAGTTTGGCCCAATGCCAAGTACGTAATGA 853
QY 695 AAACAGACTCAGACATGTTTCATCACTACTGGCAATTTAGTGAAGTATCTTTTAAA 908

Search completed: April 12, 2003, 04:34:49
Job time : 54.0484 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:22:18 ; Search time 1024.39 Seconds
(without alignments)
15994.713 Million cell updates/sec

Title: US-09-914-152-3_COPY_10000_10562

Perfect score: 563

Sequence: 1 attatcatcccacactgtc.....aaatgttggaactaactctt 563

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pi.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	563	100.0	2762	9	AB020337	AB020337 Homo sapi
2	563	100.0	2775	6	E38419	E38419 Novel polyyp
3	563	100.0	10562	6	E38420	E38420 Novel polyyp
4	563	100.0	170121	9	AF064860	AF064860 Homo sapi
5	563	100.0	340000	9	HS21C080	AL163280 Homo sapi
6	41	7.3	42614	3	CEB0513	Z82256 Caenorhabdi
7	40.4	7.2	2207	2	AC020843	AC020843 Mus muscu
c 8	40.4	7.2	175861	2	AC120346	AC120346 Mus muscu
9	40.4	7.2	196900	2	AC020851	AC020851 Mus muscu
c 10	40	7.1	125020	9	AF429315	AF429315 Homo sapi
c 11	40	7.1	173405	2	AC125587	AC125587 Rattus no
12	39.8	7.1	126119	2	AC127018	AC127018 Medicago
13	39.2	7.0	175410	2	AC102593	AC102593 Mus muscu
14	39.2	7.0	185674	2	AC010177	AC010177 Homo sapi
c 15	39.2	7.0	210010	9	AC069240	AC069240 Homo sapi
c 16	39	6.9	156168	2	AC106285	AC106285 Rattus no
c 17	38.8	6.9	57464	9	AL445530	AL445530 Human DNA
c 18	38.8	6.9	125020	9	AF429315	AF429315 Homo sapi
c 19	38.6	6.9	129007	9	AL627133	AL627133 Human DNA
c 20	38.6	6.9	146366	9	AL627231	AL627231 Human DNA
c 21	38.6	6.9	170466	2	AC129242	AC129242 Rattus no
c 22	38.6	6.9	193320	9	AC010105	AC010105 Homo sapi
c 23	38.4	6.8	145658	2	AC128174	AC128174 Rattus no
c 24	38.4	6.8	170220	2	AC120333	AC120333 Rattus no
c 25	38.4	6.8	177076	2	AC092949	AC092949 Homo sapi
c 26	38.4	6.8	177127	2	AC026287	AC026287 Homo sapi
c 27	38.4	6.8	178197	2	AC094124	AC094124 Rattus no
c 28	38.4	6.8	183357	9	AC068759	AC068759 Homo sapi
c 29	38.4	6.8	216392	2	AC095445	AC095445 Rattus no
c 30	38.2	6.8	80250	5	AL691516	AL691516 Zebrafish
c 31	38.2	6.8	147227	8	OSTN00013	AL606448 Oryza sat
c 32	38.2	6.8	188631	2	AC093161	AC093161 Rattus no
c 33	38	6.7	49863	2	AC068328	AC068328 Homo sapi
c 34	38	6.7	54454	2	AL596328	Continuation (5 of
c 35	38	6.7	164443	10	AL672285	AL672285 Mouse DNA
c 36	38	6.7	173949	2	AC027217	AC027217 Homo sapi
c 37	38	6.7	177698	2	AC091756	AC091756 Sus scrof
c 38	38	6.7	178323	9	AL354810	AL354810 Human DNA
c 39	38	6.7	207865	4	AC091755	AC091755 Sus scrof
c 40	37.8	6.7	95746	2	AC097111	AC097111 Oryza sat
c 41	37.8	6.7	146261	2	AC097112	AC097112 Oryza sat
c 42	37.8	6.7	235738	2	AC121365	AC121365 Oryza sat
c 43	37.6	6.7	7218	6	I66494	I66494 Sequence 14
c 44	37.6	6.7	39999	8	SPBP23A10	AL136535 S.pombe c
c 45	37.6	6.7	110785	9	AC074183	AC074183 Homo sapi

ALIGNMENTS

RESULT 1	AB020337	2762 bp	mrna	linear	PRI 14-APR-2000
LOCUS	Homo sapiens mRNA for UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5, complete cds.				
DEFINITION	AB020337				
ACCESSION	AB020337.1	GI:4835502			
VERSION	UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5.				
KEYWORDS	Homo sapiens Adenocarcinoma cell_line:Colo 205 CDNA to mRNA.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (sites)				
AUTHORS	Ishihara, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M., Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T. and				

Pred. No. is the number of results predicted by chance to have a

Narimatsu,H.
Cloning, expression, and characterization of a novel
UDP-galactose:beta-N-acetylglucosamine
betal,3-galactosyltransferase (beta3Gal-T5) responsible for
synthesis of type I chain in colorectal and pancreatic epithelia
and tumor cells derived therefrom
J. Biol. Chem. 274 (18), 12499-12507 (1999)
99230269
Narimatsu,H.
Isshiki,S., Togayachi,A. and Narimatsu,H.
Direct Submission
Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University,
Institute of Life Science, 1-236, Tangi-cho, Hachioji, Tokyo
192-8577, Japan (E-mail: sissiki@ipo.iijnet.or.jp,
Tel:81-426-91-9466, Fax:81-426-91-9315)
Location/Qualifiers
1. .2762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/cell_line="Colo 205"
/cell_type="Adenocarcinoma"
1. .2762
/gene="betal,3-Galt 5"
1. .273
/gene="betal,3-Galt 5"
/number=1
274. .433
/gene="betal,3-Galt 5"
/number=3
434. .2762
/gene="betal,3-Galt 5"
/number=4
434. .1366
/gene="betal,3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BA77664.1"
/db_xref="GI:4835503"
/translation="MAPFKRLMYICLLVLGAICLIFSMVSLNPFKEQSVYKKGNF
LKLPTDCRQPPFLVLLVTSKQLAERMAIRQWGMKRWKGLKFTFLIGTSS
AAETKEVDSQRIQKDFLDVYNLTKTMGIEWHVFPCQAFAFMKTDSDMF
INVDITELLKKNRTTFRTGFLKNEFIRPFKSWFVSKSEYPDWRYPPFCSTG
YVFGDASOVYNSKSVYIKLEDFVGLCLERLNLRLBELHSQTFPPFGGLRFVSC
LFRIVACHFKPRLLDYWALENSRGDCPPV"
2762
/gene="betal,3-Galt 5"
/note="45 a nucleotides"
polyA_site
BASE COUNT 651 a 703 c 679 g 729 t
ORIGIN
Query Match 100.0%; Score 563; DB 9; Length 2762;
Best Local Similarity 100.0%; Pred. No. 9.9e-152;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTTCATACCCCAACTGTCCTTGTGTTTGTATCAATGGGACCACTGCCCCAGGA 60
Db 2200 ATTTCATACCCCAACTGTCCTTGTGTTTGTATCAATGGGACCACTGCCCCAGGA 2259
QY 61 GCACATTAGGCTCTAGTTCACACTGAGGACACTGAACCTCAGATGGGCTTCATGTGG 120
Db 2260 GCACATTAGGCTCTAGTTCACACTGAGGACACTGAACCTCAGATGGGCTTCATGTGG 2319
QY 121 GATTCTGGAGCTTCTGGGAATTCAGTTGGAGTCAAGTCAAGGATGCTCTCAAGACCCC 180
Db 2320 GATTCTGGAGCTTCTGGGAATTCAGTTGGAGTCAAGTCAAGGATGCTCTCAAGACCCC 2379
QY 181 TCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCTCGTCCACTTCCC 240
Db 2380 TCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCTCGTCCACTTCCC 2439
QY 241 AAGCCTGAGCAAGCTCATCTTCATTGAATGCTCATCTTTGGCCGAGGAACAACCTGAAC 300

Db 2440 AAGCCTGAGCAAGCTCATCTTCATTGAATGCTCATCTTTGGCCGAGGAACAACCTGAAC 2499
QY 301 TGTGGTTTGTCTTACGCTTCAGTTTGTCTCGCTGCCCTCCCTACCCAGAGGTTTGTGCGA 360
Db 2500 TGTGGTTTGTCTTACGCTTCAGTTTGTCTCGCTGCCCTCCCTACCCAGAGGTTTGTGCGA 2559
QY 361 GCCTGTGTTGTCAGGGTTGTATAAACCAAGGACTTCTGTTAGTTTGGCCATTGAGCCAT 420
Db 2560 GCCTGTGTTGTCAGGGTTGTATAAACCAAGGACTTCTGTTAGTTTGGCCATTGAGCCAT 2619
QY 421 GGTCACTGACATGCAAGTAATCTTGTCTCTTAATATAGAATGATTTTCTTTAAT 480
Db 2620 GGTCACTGACATGCAAGTAATCTTGTCTCTTAATATAGAATGATTTTCTTTAAT 2679
QY 481 TTTTACTTTACAGACATTTACTTTGTACTCAGAGAGAGGCTTCACATGCTGTGTGACA 540
Db 2680 TTTTACTTTACAGACATTTACTTTGTACTCAGAGAGAGGCTTCACATGCTGTGTGACA 2739
QY 541 TATAAATGTTGGACTAAACTCTT 563
Db 2740 TATAAATGTTGGACTAAACTCTT 2762
RESULT 2
E38419
LOCUS Novel polypeptide. 2775 bp DNA linear PAT 31-JAN-2002
DEFINITION E38419
ACCESSION E38419
VERSION E38419.1 GI:18626993
KEYWORDS JP 2000245464-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Narimatsu,H., Isshiki,S., Togayachi,A. and Sasaki,K.
TITLE Novel polypeptide
JOURNAL Patent: JP 2000245464-A 1 12-SEP-2000;
COMMENT KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464-A/1
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYACHI, PI
KATSUTOSHI SASAKI
PC C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12N9/10,C12P19/00, PC
C12P21/02,
PC C12P21/08,C12Q1/68,G01N33/53//C12N1/21,C12N1/185), (C12N5/10,
C12R1:91),
PC (C12P21/02,C12R1:185), (C12P21/02,C12R1:91),
PC (C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS (402)..(1331).
FEATURES
source
1. .2775
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 691 a 698 c 669 g 727 t
ORIGIN
Query Match 100.0%; Score 563; DB 6; Length 2775;
Best Local Similarity 100.0%; Pred. No. 9.9e-152;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTTCATACCCCAACTGTCCTTGTGTTTGTATCAATGGGACCACTGCCCCAGGA 60
Db 2168 ATTTCATACCCCAACTGTCCTTGTGTTTGTATCAATGGGACCACTGCCCCAGGA 2227
QY 61 GCACATTAGGCTCTCAGTTCACACTGAAGACACTTGAACCTCAGATGGGCTTCATGTGG 120

Reichelt, J., Kauer, G., Bloecker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrsach, H., Reinhardt, R. and Yaspo, M. Laure.

The DNA sequence of human chromosome 21

Nature 405 (6784), 311-319 (2000)

20289799

10830953

2 (bases 1 to 170121)

Rump, A., Dagand, E., Hildmann, T., Nordsiek, G., Drescher, B., Weber, J., Schattevoy, R., Yaspo, M.-L. and Rosenthal, A.

Direct Submission

Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

3 (bases 1 to 170121)

Rump, A., Dagand, E., Hildmann, T., Nordsiek, G., Drescher, B., Weber, J., Schattevoy, R., Yaspo, M.-L., Rosenthal, A., Yaspo, M.-L. and Rosenthal, A.

Direct Submission

Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

On Feb 27, 2002 this sequence version replaced gi:3171153.

Location/Qualifiers

1. 170121

/organism="Homo sapiens"

/db.xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="PAC 70124"

BASE COUNT 45184 a 36756 c 37509 g 50572 t

ORIGIN

Query Match 100.0%; Score 563; DB 9; Length 170121;

Best Local Similarity 100.0%; Pred. No. 1.3e-151;

Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTCATCACCCCAACTGTCCTGTTTTCATCAATGGGACAGCCACTGCCCCAGGA 60

DB 91814 ATTTCATCACCCCAACTGTCCTGTTTTCATCAATGGGACAGCCACTGCCCCAGGA 91873

QY 61 GCACATTTAGGGCTCTCAGTTCACACTGAAGCAGAGTTGAAGTGGGGTTCATGTGG 120

DB 91874 GCACATTTAGGGCTCTCAGTTCACACTGAAGCAGAGTTGAAGTGGGGTTCATGTGG 91933

QY 121 GATTCGGGAGCTTCTCGGGAATTCAGTTGGAGTCAAGTCAAGTGCCTCAAGGACCCC 180

DB 91934 GATTCGGGAGCTTCTCGGGAATTCAGTTGGAGTCAAGTCAAGTGCCTCAAGGACCCC 91993

QY 181 TCGGCTCAGAGCCCTAAAGTGGCCCTGNGAGAGGGTGTCTCGCTCCACTTCCC 240

DB 91994 TCGGCTCAGAGCCCTAAAGTGGCCCTGNGAGAGGGTGTCTCGCTCCACTTCCC 92053

QY 241 AGCCTGAGCCCAAGCTCATCTTCATTGAATGCTCATTTGGCCGAGGAACAACCTGAACCT 300

DB 92054 AGCCTGAGCCCAAGCTCATCTTCATTGAATGCTCATTTGGCCGAGGAACAACCTGAACCT 92113

QY 301 TGTGTTTGTCTGTAGCTTCAGTTGCTCCGCTGCTCTCACTCCAGAGGTTTGTGCGA 360

DB 92114 TGTGTTTGTCTGTAGCTTCAGTTGCTCCGCTGCTCTCACTCCAGAGGTTTGTGCGA 92173

QY 361 GCCTGTGTGAGGGTTGTATAAACCAAGGTAAGTTCGTTAGTTTGGCCCATTCAGCCAT 420

DB 92174 GCCTGTGTGAGGGTTGTATAAACCAAGGTAAGTTCGTTAGTTTGGCCCATTCAGCCAT 92233

QY 421 GGTACGTCGACATCAAGTAAGTTCCTCTCTTAATATAGAAATGATTTTCTTTTAAT 480

DB 92234 GGTACGTCGACATCAAGTAAGTTCCTCTCTTAATATAGAAATGATTTTCTTTTAAT 92293

QY 481 TTTTACTTTACCAGACTTTACTTTGTACTCAGAGAGGCGCTCAGCTGGCTGTGTCA 540

DB 92294 TTTTACTTTACCAGACTTTACTTTGTACTCAGAGAGGCGCTCAGCTGGCTGTGTCA 92353

QY 541 TATAAATGTTGGACTAAACTTT 563

DB 92354 TATAAATGTTGGACTAAACTTT 92376

RESULT 5

HS21C080

LOCUS

DEFINITION

AL163280 AP001735 BA000005

ACCESSION

AL163280.2 GI:7717369

VERSION

340000 bp DNA linear PRI 24-MAY-2000

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 340000)

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakai, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornisch, K., Brandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloecker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrsach, H., Reinhardt, R. and Yaspo, M.L.

Direct Submission

Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

The Chromosome 21 Mapping and Sequencing Consortium consists of

* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,

* e.mail: sakaki@gscc.riken.go.jp

* URL: http://hgp.gsc.riken.go.jp/

and

* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,

* e.mail: gscj-submit@genome.imb-jena.de

* URL: http://genome.imb-jena.de/

and

* Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8582, Japan,

* e.mail: shimizu@mb.med.keio.ac.jp

* URL: http://adenine.dmb.med.keio.ac.jp/

and

* GBF, Dept. of Genome Analysis,

* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de

* URL: http://genome.gbf.de/

and

* Max-Planck Institute for Molecular Genetics,

* Ihnestr. 73, D-14195 Berlin, Germany,

* e.mail: info-chr21@molgen.mpg.de

* URL: http://chr21.rz-berlin.mpg.de/

Location/Qualifiers

1. 340000

/organism="Homo sapiens"

/db.xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

FEATURES

source

```
source
<1. 125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_lib="P160L9, 5' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121897"
73410..243533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF064860"
224137..>340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121782"
286628..>340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088..1556
/note="LIM4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398..2495
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
complement(2496..2878)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region
2879..2931
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
2932..3004
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
3005..3074
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
gene
3473..19157
/feature="SH3BGR"
join(<3473..3565,12649..12678,15331..15460,18739..19157)
mRNA
/feature="SH3BGR"
join(3473..3565,12649..12678,15331..15426)
CDS
/partial
/feature="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/db_xref="GI:7717370"
/translation="GSEKAEGGTEAQKESVDVGNLPEAQEKNEEGGTATEETEE
IAMEGAGEEEETAEAGEEPGEDEDS"
3473..3565
/feature="SH3BGR"
/number=4
exon
```

```
repeat_region
5542..5829
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6887..7067)
/note="LIMD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8482..8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8865..8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
9131..9424
/note="AluSq"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(9723..9882)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9989..10077)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10078..10429)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
10980..11000
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(11002..11282)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/feature="SH3BGR"
/number=5
complement(13115..13425)
/note="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
/note="(CA)n"
/rpt_family="Simple_repeat"
```

```
Query Match 100.0%; Score 563; DB 9; Length 340000;
Best Local Similarity 100.0%; Pred. No. 1.4e-151;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTTCATCACCCCACTGTCCTGTGTTTTGATCATGGGACGACGACCTCCCCAGGA 60
|||||
```


TKDAGSTPAKDFLTKTLTHFLKWMKVPLLASFLKKVBEIEKLGDIETQNYMIGDVNQ
VRDLVFLVSS"
complement(join(10060..10224,10273..10359,10906..10998,
14802..14807))
/gene="B0513.9"
complement(join(10060..10224,10273..10359,10906..10998,
14802..14807))
/gene="B0513.9"
/note="cDNA EST EMBL:AU110342 comes from this gene
cDNA EST EMBL:AU114094 comes from this gene"
/codon_start=1
/protein_id="CAC70077.1"
/db_xref="GI:15718043"
/translation="MMGMENDSNKPSNDQOKKMKQLNRQVDSVRKAVMVDNVERILE
RGLRDNIERSEQLNATSANFKTARKVQKFCMLNAKWTIVLSILVILVEFVLLLL
ILHTGVIGKRNND"
complement(join(18399..18611,19636..19836,19964..20168,
20467..20807,21444..21699,22537..22640,23569..23943))
/gene="B0513.5"
complement(join(18399..18611,19636..19836,19964..20168,
20467..20807,21444..21699,22537..22640,23569..23943))
/gene="B0513.5"
/note="contains similarity to Pfam domain: PF01619
(Proline dehydrogenase), Score=585.9, E-value=8.2e-173,
N=1
cDNA EST Yk4f12.5 comes from this gene
cDNA EST Yk50a10.5 comes from this gene
cDNA EST Yk70g4.5 comes from this gene
cDNA EST Yk170a11.5 comes from this gene
cDNA EST Yk170a11.3 comes from this gene
cDNA EST Yk637a8.3 comes from this gene
cDNA EST Yk637a8.5 comes from this gene
cDNA EST Yk248f7.3 comes from this gene
cDNA EST Yk248f7.5 comes from this gene
cDNA EST Yk277a12.5 comes from this gene"
/codon_start=1
/protein_id="CAB05117.1"
/db_xref="GI:3873847"
/db_xref="SWISS-PROT:O45228"
/translation="MKIPVALVTIIIEFFQSKSNTELVRALVLRCLGTLQVLRNQG
ILNMRVLGKNTFKTKLNTFFHFAGETEEVHRVHVKLVKSILDSVEA
DITSEATDKTKVTSVATVPKPAATPVDAKTLETRERTVHEFGDRGROVSAR
TYFVEGECDDNRDIPKDSINAVASATNEGFVAVKIITLGRPQLLLKLSRAIVGTQ
NFFKALGSGSLQDRITQSFEYKRLGELGVKIDTSVKFFDEVDSDGIVDLHG
NHLDDHVQGLQFVLNKTGSLLEQNLNNEDEEFNNMVRRLDVAEIAIEKV
RIMVDAQTLQPAISKITIEMKKYNKGNFTYQAYLKGTLQNMADMQVARR
GWHFGAKLVGRATYVRFALNLMKEKISPSERVMAQLYGMCDQVGFSLGQAFSVYK
NVSVMASHNEDTVRFALNLMKEKISPSERVMAQLYGMCDQVGFSLGQAFSVYK
YLPYGPVEEVLPLSLRALENGSVLKANKERDLLKELKRISSEFFKARSSSS"
19876..19947
/gene="B0513.tl"
19876..19947
/gene="B0513.tl"
/product="trna-His"
/note="GTG His H-trna
predicted using tRNAscan-SE-1.11
preliminary prediction
similar to trna-His"
join(29767..29893,30740..30793,30864..30967,31020..31097)
/gene="B0513.4"
join(29767..29893,30740..30793,30864..30967,31020..31097)
/gene="B0513.4"
/note="cDNA EST Yk124h11.3 comes from this gene
cDNA EST Yk309c11.3 comes from this gene
cDNA EST Yk412g9.3 comes from this gene
cDNA EST Yk438e7.5 comes from this gene
cDNA EST Yk442d6.5 comes from this gene
cDNA EST Yk388g4.3 comes from this gene
cDNA EST Yk539e12.3 comes from this gene
cDNA EST Yk544b1.3 comes from this gene
cDNA EST Yk545g8.3 comes from this gene
cDNA EST Yk547d2.3 comes from this gene
cDNA EST Yk548d7.5 comes from this gene"

/codon_start=1
/protein_id="CAB05116.1"
/db_xref="GI:3873846"
/db_xref="SPTREMBL:O45227"
/translation="MNSLILISIAALAVVHAQYGAQAPAYAPAYQYPPAALNDNY
VCTIQANTGLFGQGGKHYRPTVNYCQDTYKNDSCKCKCKMAARIQGTNINEASIIIGEN
MYLEKQPLVCVCSNPTSY"
join(31853..31889,32044..32195)
/gene="B0513.3"
join(31853..31889,32044..32195)
/gene="B0513.3"
/note="contains similarity to Pfam domain: PF01779
(Ribosomal L29e protein family), Score=76.3,
E-value=2e-19, N=1
cDNA EST Yk493f9.3 comes from this gene
cDNA EST Yk493f9.5 comes from this gene"
/codon_start=1
/protein_id="CAB05115.1"
/db_xref="GI:3873845"
/db_xref="SPTREMBL:O45226"
/translation="MAKSKNHTNHQNKKAHRNGITPKKKHIFLSMKGVDAKFINKLR
FSRKNKXQINKSVESKA"
Query Match 7.3%; Score 41; DB 3; Length 42614;
Best Local Similarity 57.4%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 372 AGGGTTGTATATAACCAAGTACTTCGTTAGTTTTGCCCATTCACCAATGTCACGTGAC 431
Db 16079 AGGAAAGCAGAAATAATTCACCTACTCAGTTTTGGCACTCATTAATTTGGGAAGTCGA 16138
QY 432 ATGCAAGTAATCTGCTCTCTAATATAGAAATGATTTCTTTTAAATTTTACTTTAC 491
Db 16139 CTCAAAAGGTGTTTACACTACTTTTACACTTCAATTTATTTTAAATTTTGTGTTAA 16198
QY 492 CAGACTTTA 500
Db 16199 AACATTCTA 16207
RESULT 7
AC020843 Mus musculus clone RP21-186E16, LOW-PASS SEQUENCE SAMPLING.
AC020843 HTG: HTGS_PHASE0.
AC020843.1 GI:6686465
AC020843 Mus musculus.
AC020843 Mus musculus.
AC020843 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AC020843 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AC020843 DOE Joint Genome Institute.
AC020843 Sequencing of Mouse
AC020843 Unpublished
AC020843 2 (bases 1 to 2207)
AC020843 DOE Joint Genome Institute.
AC020843 Direct Submission
AC020843 Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
AC020843 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AC020843 * NOTE: This record contains 2 individual
AC020843 * sequencing reads that have not been assembled into
AC020843 * contigs. Runs of N are used to separate the reads
AC020843 * and the order in which they appear is completely
AC020843 * arbitrary. Low-pass sequence sampling is useful for
AC020843 * identifying clones that may be gene-rich and allows
AC020843 * overlap relationships among clones to be deduced.
AC020843 * However, it should not be assumed that this clone
AC020843 * will be sequenced to completion. In the event that
AC020843 * the record is updated, the accession number will
AC020843 * be preserved.
AC020843 * 1 485: contig of 485 bp in length
AC020843 * gap of unknown length

FEATURES	source
location/Qualifiers	
1. :175861...	
DEFINITION	
Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33 unordered pieces.	

```
misc_feature      1. .28835
```

```
misc_feature      28936.29562
```

```
misc_feature 296663. 30328
/note=assembly_fragment
```

```
misc feature 30429. 32005
/note=assembly_fragment
```

```
misc feature 32106. 33128
/note=assembly_fragment
```

```
misc feature 33229 34512
/note="assembly_fragment
```

```
misc feature 34613 36079
/note="assembly_fragment
```

```
misc feature 36180 37483
/note="assembly_fragment
```

```
misc feature 37584 39166
/note="assembly_fragment
```

```
misc feature 39267 41065
/note="assembly_fragment
```

```
misc feature 41166 43202
/note="assembly_fragment
```

```
misc feature 43204 47480
/note="assembly_fragment"
```

```
misc feature 47500 50504
/note="assembly_fragment"
```

```
Feature    50705  55370  
/note="assembly_fragment"
```

```

/note="assembly_fragment"
FC471 C4070

```

```

/note="assembly_fragment"
CA100 70010

```

```

...
/note="assembly_fragment"
7270 0150
...

```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```

/note="assembly_fragment

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

BASE COUNT   48903 a 39540 c 39029 g 46286 t 2103 others
ORIGIN
Query Match          7.2% Score 40.4; DB 2: Length 175861;
Best Local Similarity 61.3%; Pred.No.1.9;
Matches 65; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY    35 AATGGGACCAAGCAGCCGCCAGGACACTTTTAGGGCTCTCACGTTCAAACTGAAGACA 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    61120 ATTAGTCACTATTGCAGCCCTCGGGGACACTTCAGACCTCTTAAGTTCAAACCTCAAGGTA 61061
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    95 GTTGAACTCAGATGGGTTTCATGTGGGATTTCTGGAGCTTTCTGGG 140
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    61060 GCCAGGTACAGGCATGCTGCTGTATGGATACC CGGGAGCTTTCCTGG 61015
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AC020851
LOCUS       Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
DEFINITION unorderd pieces.
ACCESSION  AC020851
VERSION     AC020851.2 GI:9211212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 196900)
AUTHORS    DOE Joint Genome Institute.
TITLE       Sequencing of Mouse
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 196900)
AUTHORS    DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:5686457.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
Project Information
Center Project Name: 1437240
Center clone name: RPCI-21_467L12
            -----
Summary Statistics
Consensus quality: 164618 bases at least Q40
Consensus quality: 180867 bases at least Q30
Consensus quality: 183777 bases at least Q20
Estimated insert size: 147000; pulse field gel estimation
Estimated insert size: 193700; sum-of-contigs estimation
Quality coverage: 5.67 in Q20 bases; pulse field gel estimation
Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1541: contig of 1541 bp in length
* 1542 1641: gap of unknown length
* 1642 3492: contig of 1851 bp in length
* 3493 3592: gap of unknown length
* 3593 4738: contig of 1146 bp in length
* 4739 4838: gap of unknown length
* 4839 7248: contig of 2410 bp in length
* 7249 7348: gap of unknown length
* 7349 91978: contig of 1830 bp in length

```



```

* 44153 51366: contig of 7214 bp in length
* 51367 51466: gap of unknown length
* 51467 61504: contig of 10038 bp in length
* 61505 61604: gap of unknown length
* 61605 71420: contig of 9816 bp in length
* 71421 86316: contig of 14796 bp in length
* 86317 86416: gap of unknown length
* 86417 99317: contig of 12901 bp in length
* 99318 99417: gap of unknown length
* 99418 126119: contig of 26702 bp in length.
FEATURES
    source
        location/Qualifiers
            1..126119
            /organism="Medicago truncatula"
            /db_xref="taxon:3880"
            /clone="mth1-8a13"
            /clone_lib="Medicago truncatula BAC library H1"
BASE COUNT 41110 a 21324 c 21213 g 41128 t 1344 others
ORIGIN
Query Match 7.1%; Score 39.8; DB 2; Length 126119;
Best Local Similarity 55.4%; Pred. No. 2.8;
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 352 TTGTGCGGAGCTGTGTTGCAGGTTGTATATAAACCAAGGACCTTCGTTAGTTTGGCCCA 411
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117583 TTGTGCGTGGTGGAGTTTCGATTCGCCAACCTTACATATATATATGTCATGTTTCATATCAA 117642
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 412 TTCAGGCATGGTCACGTGCATGCAAGTAATCTTGCCTCCCTAATATAGAAATGATTTTT 471
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117643 TTCAGTAAGTTCACGTGCACTTCRAATATATTAGTTTGAAGATAGTTTATTTTATTTG 117702
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 472 CTTTTTAATTTTACTTTA 490
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117703 AAAATATTTTATATTTTA 117721
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13

```

AC102593 175410 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP23-341M12, WORKING DRAFT SEQUENCE, 15
DEFINITION unordered pieces.
ACCESSION AC102593
VERSION AC102593.2 GI:22381580
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 175410)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-341M12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175410)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,D.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testa,S., Theodore,J.,

```

```

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 175410)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testa,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061679.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19073
Center clone name: 341_M12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 171538 bases at least Q40
Consensus quality: 173037 bases at least Q30
Consensus quality: 173658 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 174010; sum-of-contigs
Quality coverage: 7.8 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 9748 9847: contig of 9747 bp in length
* 9848 92714: gap of 100 bp
* 92715 92814: contig of 82867 bp in length
* 92815 95450: contig of 2636 bp in length
* 95451 98550: gap of 100 bp
* 95551 98272: contig of 2722 bp in length
* 98273 98372: gap of 100 bp
* 98373 101914: contig of 3542 bp in length
* 101915 102014: gap of 100 bp
* 102015 104240: contig of 2226 bp in length
* 104241 104340: gap of 100 bp
* 104341 108050: contig of 3710 bp in length

```

Center: Baylor Col

School of Medicine

sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES
Source Location/Qualifiers
1..210010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-385N17"

misc_feature
1..2000

/note="overlaps bases 97728..99727 of clone AC009260"
/function="clone overlap"

repeat_region
1519..1553

/rpt_family="(TC)n"

repeat_region
1656..1707

/rpt_family="CT-rich"

repeat_region
3921..3944

/rpt_family="(TTA)n"

STS
5374..5538

/standard_name="98275"

STS
8015..8215

/standard_name="77808"

STS
8018..8146

/standard_name="97556"

repeat_region
10803..10884

/rpt_family="(TG)n"

repeat_region
12107..12151

/rpt_family="(TA)n"

repeat_region
12324..12463

/rpt_family="(CA)n"

repeat_region
13694..13857

/rpt_family="FRAM"

repeat_region
13860..14529

/rpt_family="L2"

repeat_region
14180..14190

/rpt_family="MIR"

repeat_region
14191..14493

/rpt_family="L1PA10"

repeat_region
14494..14628

/rpt_family="MIR"

repeat_region
14639..14790

/rpt_family="L2"

repeat_region
complement(14851..14938)

repeat_region
complement(14949..15001)
/rpt_family="MER47"
repeat_region
15683..15708
/rpt_family="AT-rich"
repeat_region
16108..16295
/rpt_family="MIR"
repeat_region
16296..16607
/rpt_family="MER7A"
repeat_region
16608..16629
/rpt_family="MIR"
repeat_region
17336..18010
/rpt_family="L1MA2"
repeat_region
18202..18524
/rpt_family="MLT1H"
repeat_region
18592..18901
/rpt_family="AluY"
repeat_region
19222..19441
/rpt_family="MLT1K"
repeat_region
19449..19592
/rpt_family="MIR"
repeat_region
19732..19927
/rpt_family="MIR"
repeat_region
21192..21339
/rpt_family="MIR"
repeat_region
complement(21895..22006)
/rpt_family="L2"
repeat_region
complement(22205..22353)
/rpt_family="MIR"
repeat_region
22879..23241
/rpt_family="L1MC4"
repeat_region
23349..23479
/rpt_family="MLT1K"
repeat_region
complement(23555..23663)
/rpt_family="L2"
repeat_region
23665..24018
/rpt_family="MER46B"

Query Match

Best Local Similarity 7.0%; Score 39.2; DB 9; Length 210010;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 398 GTAGTTTGGCCATTGACCCATGTCACGTGACATGCAAAAGTAATCTGCTCAATTA 457

Db 129257 GATTGGTTTTCCTCCTCAGTTAAGTTCCCTTGAATTCCTGCAAAATGTTATTCATCA 129198

QY 458 TAGAAATGATTTTCTTTTAAATTTTACTTTTACAGACTTTTGTCTACTCAGA 513

Db 129197 GTGCTTATTTTCCTTTTCGTTTTCCTTTTCTTTTCTTTTCTTTTGGAGA 129142

Search completed: April 11, 2003, 21:30:31
Job time : 1922.39 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:50:38 ; Search time 93.5557 Seconds
(without alignments)
13552.085 Million cell updates/sec

Title: US-09-914-152-3_COPY_10000_10562

Perfect score: 563

Sequence: 1 atttcataccccaactgtc.....aaatgttggaactaaacttt 563

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	563	100.0	2775	21	AAA93875 Human beta-1,3 gal
2	563	100.0	10562	21	AAA93876 Human beta3gal-ts
3	60	10.7	60	24	ABN47639 Human spliced tran
4	36.6	6.5	706	22	AAH04137 Human cDNA clone (
5	36.6	6.5	2208	22	AAH15492 Human cDNA sequenc
6	36	6.4	16439	24	ABL32887 Human immune syste
7	35.8	6.4	1775	22	AAH45194 Chrysanthemum mori
8	35.8	6.4	2447	23	ABJ08300 Drosophila melanog
9	35.6	6.3	11280	24	ABN60004 Human chemically m

10	35	6.2	21001	23	ABL07952	Drosophila melanog
11	34.8	6.2	30143	23	ABL18664	Drosophila melanog
12	34.8	6.2	122888	24	ABK3569	Human cDNA differe
13	34.6	6.1	521	22	ABA61695	Human foetal liver
14	34.6	6.1	521	22	AAK09997	Human brain expres
15	34.6	6.1	521	22	AAK35891	Human bone marrow
16	34.6	6.1	521	22	AAK1607	Probe #10293 used
17	34.6	6.1	7908	22	AAK66239	Human immune/haema
18	34.4	6.1	564	22	ABA63995	Human foetal liver
19	34.4	6.1	564	22	ABA76192	Human foetal liver
20	34.4	6.1	564	22	AAK12495	Human brain expres
21	34.4	6.1	564	22	AAK24854	Human brain expres
22	34.4	6.1	564	22	AAK38205	Human bone marrow
23	34.4	6.1	564	22	AAK50847	Human bone marrow
24	34.4	6.1	564	22	AAI18996	Probe #8929 for ge
25	34.4	6.1	564	22	AAI27878	Probe #17811 for g
26	34.4	6.1	564	22	AAI44133	Probe #12819 used
27	34.4	6.1	564	22	AAI56860	Probe #25546 used
28	34.4	6.1	564	24	ABS12240	Human genome-deriv
29	34.4	6.1	564	24	ABS24348	Human genome-deriv
30	34.4	6.1	5275	24	ABS24284	Human immune syste
31	34.2	6.1	1074	24	ABN70373	Streptococcus poly
32	34.2	6.1	1161	24	ABN67231	Streptococcus poly
33	34.2	6.1	12123	22	AAS99855	Genomic sequence #
34	34.2	6.1	12123	22	AAK90226	Human digestive sy
35	34	6.0	5096	23	ABL30415	Drosophila melanog
36	34	6.0	15966	23	ABL30414	Drosophila melanog
37	33.8	6.0	16682	22	AAK70010	Human immune/haema
38	33.8	6.0	16682	22	AAK97998	Human immune/haema
39	33.8	6.0	21436	22	AAK70011	Human immune/haema
40	33.8	6.0	21436	22	AAK97999	Human immune/haema
41	33.6	6.0	1493	22	AAK67730	Human immune/haema
42	33.4	5.9	436	23	ABV50237	Human prostate exp
43	33.4	5.9	5914	24	ABL34168	Human immune syste
44	33.4	5.9	8592	24	ABL33983	Human immune syste
45	33.4	5.9	8711	22	AAS46700	Tumour suppressor

ALIGNMENTS

RESULT 1
AAA93875
ID AAA93875 standard; DNA; 2775 BP.
AC AAA93875;
XX
XX
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
XX digestive system; ds.
XX
OS Homo sapiens.
XX
PN WO2000050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI: 2000-549409/50.
XX
DR P-PSDB; AAB93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
XX synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of

PT digestive system cancer -
XX
PS Claim 5; Page 99-102; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylgalucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents Beta-1,3 galactose
CC transferase encoding DNA.
XX
XX Sequence 2775 BP; 681 A; 698 G; 669 G; 727 T; 0 other;
XX
Query Match 100.0%; Score 563; DB 21; Length 2775;
Best Local Similarity 100.0%; Pred. No. 2.9e-159;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTTCAATCACCCCAACTGTCCTCTGTTTGTGATCAATGGGACAGCCACTGCCCCAGGA 60
DB 2168 ATTTCAATCACCCCAACTGTCCTCTGTTTGTGATCAATGGGACAGCCACTGCCCCAGGA 2227
QY 61 GCATTTAGGGCTCTCAGTTCAAACTGAAGCAGAGTTGAATCAGATGGGGTTTCATGTGG 120
DB 2228 GCATTTAGGGCTCTCAGTTCAAACTGAAGCAGAGTTGAATCAGATGGGGTTTCATGTGG 2287
QY 121 GATTCGTGGAGCTTCTCGGAAATTCAGTTGGAGTCAAGTCAAGTGCCTCAAGGACCC 180
DB 2288 GATTCGTGGAGCTTCTCGGAAATTCAGTTGGAGTCAAGTGCCTCAAGGACCC 2347
QY 181 TCGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGAGGTGGTCTCGTCCACATCC 240
DB 2348 TCGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGAGGTGGTCTCGTCCACATCC 2407
QY 241 AAGCCTGAGCCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAATGAAT 300
DB 2408 AAGCCTGAGCCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAATGAAT 2467
QY 301 TGTGTTTGTGTTAGCCCTTACGTTTGTCTCCGCTGCTCTACCCAGAGGTTGTGCGA 360
DB 2458 TGTGTTTGTGTTAGCCCTTACGTTTGTCTCCGCTGCTCTACCCAGAGGTTGTGCGA 2527
QY 361 GCCTGTGTGTCAGGGTTGTATATAAACCAAGGTACTTCGTTAGTTTGGCCCATTCAGCCAT 420
DB 2528 GCCTGTGTGTCAGGGTTGTATATAAACCAAGGTACTTCGTTAGTTTGGCCCATTCAGCCAT 2587
QY 421 GGTACGTGACATGCAAGTAATCTTCTCTTAATATAGAAATGATTTTCTTTTAAT 480
DB 2588 GGTACGTGACATGCAAGTAATCTTCTCTTAATATAGAAATGATTTTCTTTTAAT 2647
QY 481 TTTTACTTTACGACCTTTACTTTGTACTCAGAGAAGAGCCCTCACATGCTGTGTACA 540
DB 2648 TTTTACTTTACGACCTTTACTTTGTACTCAGAGAAGAGCCCTCACATGCTGTGTACA 2707
QY 541 TATAAATGTTGGACTAAACTCTT 563
DB 2708 TATAAATGTTGGACTAAACTCTT 2730
RESULT 2
AAA93876
ID AAA93876 standard; DNA: 10562 BP.
XX
AC
XX AAA93876;
XX
DT 15-JAN-2001 (first entry)
XX

Human beta3Gal-T5 encoding DNA.
DE
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; beta3Gal-T5; ds.
XX
OS Homo sapiens.
XX
PN WO2000050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
XX WPI; 2000-549409/50.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer -
XX
PS Claim 31; Page 103-111; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylgalucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents a Beta3gal-T5
CC encoding DNA sequence.
XX
XX Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;
XX
Query Match 100.0%; Score 563; DB 21; Length 10562;
Best Local Similarity 100.0%; Pred. No. 5.2e-159;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTTCAATCACCCCAACTGTCCTCTGTTTGTGATCAATGGGACAGCCACTGCCCCAGGA 60
DB 10000 ATTTCAATCACCCCAACTGTCCTCTGTTTGTGATCAATGGGACAGCCACTGCCCCAGGA 10059
QY 61 GCATTTAGGGCTCTCAGTTCAAACTGAAGCAGAGTTGAATCAGATGGGGTTTCATGTGG 120
DB 10060 GCATTTAGGGCTCTCAGTTCAAACTGAAGCAGAGTTGAATCAGATGGGGTTTCATGTGG 10119
QY 121 GATTCGTGGAGCTTCTCGGAAATTCAGTTGGAGTCAAGTGCCTCAAGGACCC 180
DB 10120 GATTCGTGGAGCTTCTCGGAAATTCAGTTGGAGTCAAGTGCCTCAAGGACCC 10179
QY 181 TCGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGAGGTGGTCTCGTCCACTTCC 240
DB 10180 TCGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGAGGTGGTCTCGTCCACTTCC 10239
QY 241 AAGCCTGAGCCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAATGAAT 300
DB 10240 AAGCCTGAGCCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAATGAAT 10299
QY 301 TGTGTTTGTGTTAGCCCTTACGTTTGTCTCCGCTGCTCTACCCAGAGGTTGTGCGA 360
DB 10300 TGTGTTTGTGTTAGCCCTTACGTTTGTCTCCGCTGCTCTACCCAGAGGTTGTGCGA 10359
QY 361 GCCTGTGTGTCAGGGTTGTATATAAACCAAGGTACTTCGTTAGTTTGGCCCATTCAGCCAT 420

Db 10360 GCCTGTTGTCAGGGTTGTATTAARACCAAGGTACTTCGTTAGTTTGGCCATTCAGCCAT 10419
QY 421 GGTCAAGTGACATGCAAAAGTAATCTTGCTCTCTAAATATAGAAATGATTTTCTTTTAATT 480
Db 10420 GGTCACGTGACATGCAAAAGTAATCTTGCTCTCTAAATATAGAAATGATTTTCTTTTAATT 10479
QY 481 TTTTACTTTACAGACTTTTACTTTGTACTCAGAGAAGGCCTCACAATGGCTGTGTGCACA 540
Db 10480 TTTTACTTTACAGACTTTTACTTTGTACTCAGAGAAGGCCTCACAATGGCTGTGTGCACA 10539
QY 541 TATAAATGTTGACTAAACTCTT 563
Db 10540 TATAAATGTTGACTAAACTCTT 10562

RESULT 3
ABN47639
ID ABN47639 standard; DNA; 60 BP.
XX
AC ABN47639;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:20387.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
PI WPI; 2002-257383/30.
XX
DR New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
PT
XX Example 1; SEQ ID 20387; 47pp; English.
PS
XX

CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialized mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.

CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 60 BP; 13 A; 8 C; 21 G; 18 T; 0 other;
Query Match 10.7%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 2 6e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 98 GAACTCAGATGGGTTCATGTGGGATTTCTGGAGAGCTTTCTGGGAATTCAGTTGGAGTCAA 157
Db 1 GAACTCAGATGGGTTCATGTGGGATTTCTGGAGAGCTTTCTGGGAATTCAGTTGGAGTCAA 60
RESULT 4
AAH04197
ID AAH04197 standard; cDNA; 706 BP.
XX
AC AAH04197;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:1032.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 1032; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 706 BP; 251 A; 103 C; 96 G; 253 T; 3 other;

Query Match 6.5%; Score 36.6; DB 22; Length 706;
Best Local Similarity 50.9%; Pred. No. 0.9;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 380 ATAAACCAAGTACTCGTTAGTTTGGCCATTCAGCCATGTCAGTGACATGCAAG 439
DB 252 ATACGCCATATTAATGCTTCGATTTCTCATATATGATGATCAAAATTTTGGAC 311

QY 440 TAATCTTCCTCTAATATAGAAATGATTTTCTTTTAAATTTTCTTACAGACTTT 499
DB 312 TAATCTGAAGCTATTTTAAATTTATTTTAAATTTTCTGAGAAAAAATG 371

QY 500 ACTTTGTACTCAGAGAGAGCCCTCAGATGCTGTGCATATATAATGTT 550
DB 372 ATTTTCTTCTATACATGATCTAGAGAACTGATTTAAAGTTACAGCT 422

RESULT 5
AAH15492
XX ID AAH15492 standard; cDNA; 2208 BP.
XX AC AAH15492;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:13750.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13750; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

SQ Sequence 2208 BP; 690 A; 407 C; 419 G; 692 T; 0 other;

Query Match 6.5%; Score 36.6; DB 22; Length 2208;
Best Local Similarity 50.9%; Pred. No. 1.5;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 380 ATAAACCAAGTACTCGTTAGTTTGGCCATTCAGCCATGTCAGTGACATGCAAG 439
DB 252 ATACGCCATATTAATGCTTCGATTTCTCATATATGATGATCAAAATTTTGGAC 311

QY 440 TAATCTTCCTCTAATATAGAAATGATTTTCTTTTAAATTTTCTTACAGACTTT 499
DB 312 TAATCTGAAGCTATTTTAAATTTATTTTAAATTTTCTGAGAAAAAATG 371

QY 500 ACTTTGTACTCAGAGAGAGCCCTCAGATGCTGTGCATATATAATGTT 550
DB 372 ATTTTCTTCTATACATGATCTAGAGAACTGATTTAAAGTTACAGCT 422

RESULT 6
ABL32887
XX ID ABL32887 standard; DNA; 16439 BP.
XX AC ABL32887;
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 860.
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytosolic; neurotropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP07537.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 122888 BP; 28761 A; 33410 C; 31919 G; 28798 T; 0 other;
Query Match 6.2%; Score 34.8; DB 24; Length 122888;
Best Local Similarity 51.3%; Pred. No. 32;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 376 TTGATAAACCAAGGTACTTCGTTAGTTTGGCCATTCAGCCATGGTCAGTGACATGC 435
DB 967 TTTGTCAACCATATCCATGGTTTATTTGGCCATCCCTCATTTGTGAAATCTCACT 1026
QY 436 AAAGTAATCTTGCCTCCCAATATAGAATGATTTTCTTTTAAATTTTACTTTACCCAGA 495
DB 1027 TATTTCCAGTTTTCGAAGCTATAAATAATATTTGGATTATATCTTCTACAGATAG 1086
QY 496 CTTTACTTTTACTCAGAGAGGCGCTCACATGGGCTG 533
DB 1087 CTTTCCCTTTTCCATATTTAGGGGTGCTCTCTGGCTG 1124

RESULT 13
ABA61695/c
ID ABA61695 standard; DNA; 521 BP.
XX AC ABA61695;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #10000.
XX KW Human; foetal liver; gene expression: single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 1; SEQ ID NO 10000; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 521 BP; 140 A; 120 C; 104 G; 157 T; 0 other;
Query Match 6.1%; Score 34.6; DB 22; Length 521;
Best Local Similarity 50.9%; Pred. No. 3.2; Mismatches 0; Gaps 0;
Matches 82; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 388 AAGTACTTCTGTTAGTTTGGCCATTCAGCCATGGTCAGATGCAAGTAATCTTG 447
DB 458 AAGGAAGTGCATTACTCTACTATCTCCAGGAGAGTCATATGTTTCCAAACAACATG 399
QY 448 CTCTAATATAGAAATGATTTTCTTTTAAATTTTACTTTACCGACACTTACTTTGTA 507
DB 398 AGCTAATCTAGAAGTTGTTTCTTTTCAATATGAAACTTCCATTTTCCATGCTT 339
QY 508 CTCAGAGAAGAGCGCTCACATGCTGTGTCACATATAAATG 548
DB 338 GCCAGTTAGATGGAGCCAGGAGACTATAATATTAGTG 298

RESULT 14
AAK09997/c
ID AAK09997 standard; DNA; 521 BP.
XX AC AAK09997;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 9988.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS Example 4; SEQ ID NO: 9988; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX CC invention.

SQ Sequence 521 BP; 140 A; 120 C; 104 G; 157 T; 0 other;
Query Match 6.1%; Score 34.6; DB 22; Length 521;
Best Local Similarity 50.9%; Pred. No. 3.2; Mismatches 0; Gaps 0;
Matches 82; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:35:54 ; Search time 626.782 Seconds
(without alignments)
14547.424 Million cell updates/sec

Title: US-09-914-152-3_COPY_10000_10562

Perfect score: 563

Sequence: 1 atttcacaccccaactgtc.....aaatgttggaactaaactctt 563

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_eston:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	505.4	89.8	507	9	AI739518
C 2	438	77.8	455	10	AW136254
C 3	39.6	7.0	896	17	CNS0081M
C 4	38.6	6.9	742	17	CNS021SK
C 5	38.2	6.8	730	17	AZ255953
C 6	38.2	6.8	737	17	AZ986887

C	7	38.2	6.8	1201	17	CNS0161Y	AL106192 Drosophil
	8	38	6.7	1101	17	CNS0009A	AL060388 Drosophil
	9	37.8	6.7	269	13	BJ083396	BJ083396 BJO83396
	10	37.6	6.7	538	17	BH106708	BH106708 RPCI-24-4
C	11	37.4	6.6	850	17	BH56489	BH56489 BOMJ12TF
	12	37.2	6.6	1093	17	CNS01576	AL105084 Drosophil
	13	37	6.6	228	10	BBO21590	BBO21590 BBO21590
	14	37	6.6	302	10	BH364329	BH364329 BH364329
C	15	36.8	6.5	281	10	BBO60437	BBO60437 BBO60437
	16	36.6	6.5	1011	17	CNS07CYB	AL439785 T3 end of
	17	36.6	6.5	888	17	AZ327467	AZ327467 LMO050J14
	18	36.6	6.5	706	9	AU120120	AU120120 AU120120
C	19	36.6	6.5	983	12	BE733463	BE733463 601568860
	20	36.4	6.5	296	14	R63365	R63365 Y108D08.S1
C	21	36.4	6.5	482	9	AU087850	AU087850 AU087850
	22	36.4	6.5	688	17	AG171598	AG171598 Pan trogl
	23	36.4	6.5	987	17	CNS001U6	AL075445 Drosophil
	24	36.2	6.4	529	14	BM879275	BM879275 kul0e04.y
C	25	36	6.4	372	9	AI764597	AI764597 01-R-Y0-a
C	26	36	6.4	436	17	AQ180918	AQ180918 HS_3225_A
C	27	36	6.4	519	10	BH136649	BH136649 BBI36649
C	28	36	6.4	680	17	CNS03JWD	AL247414 Tetraodon
C	29	36	6.4	832	12	BG336781	BG336781 602405523
C	30	36	6.4	1101	17	CNS008NV	AL052536 Drosophil
C	31	35.8	6.4	524	17	AQ548257	AQ548257 RPCI-11-4
C	32	35.6	6.3	393	9	AL599561	AL599561 DKF2p313G
C	33	35.6	6.3	866	17	AZ675162	AZ675162 ENTMT78TR
C	34	35.4	6.3	292	10	B452574	B452574 BA452574
C	35	35.4	6.3	329	10	B5558855	B5558855 B5558855
C	36	35.4	6.3	717	17	BH586592	BH586592 BOHIM51TR
C	37	35.2	6.3	205	9	AV008665	AV008665 AV008665
C	38	35.2	6.3	381	14	C94114	C94114 C94114 Dict
C	39	35.2	6.3	430	12	BG669852	BG669852 DRNALB03
C	40	35.2	6.3	486	13	B1705278	B1705278 fr56a06.y
C	41	35.2	6.3	549	13	B1840552	B1840552 fr57f03.y
C	42	35	6.2	305	10	B201789	B201789 B201789
C	43	35	6.2	740	12	BF865202	BF865202 963057H02
C	44	35	6.2	824	17	CNS03FTP	AL242134 Tetraodon
C	45	34.8	6.2	264	12	BF424069	BF424069 sr41hil.y

ALIGNMENTS

RESULT 1
AI739518/c

LOCUS AI739518 507 bp mRNA linear EST 20-DEC-1999
DEFINITION w123a06.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:2391058 3', mRNA sequence.

ACCESSION AI739518
VERSION AI739518.1 GI:5101499
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 507)

REFERENCE

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/dbbrp/image/image.html

Insert Length: 1331 Std Error: 0.00

Seq primer: -40UP from Gibco

FEATURES	High quality sequence stop: 465.			
	Location/Qualifiers			
	1. .507			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:2391058"			
	/clone_lib="NCI_CGAP_Col6"			
	/tissue_type="colon tumor, RER"			
	/lab_host="DH108"			
	/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Col10 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo."			
BASE COUNT	154 a	122 c	115 g	116 t
ORIGIN				
	Query Match	89.8%;	Score 505.4;	DB 9; Length 507;
	Best Local Similarity	99.8%;	Pred. No. 1.2e-107;	
	Matches 506;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
QY	56	CAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACCTCAGATGGGTTCA	115	
DB	507	CAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACCTCAGATGGGTTCA	448	
QY	116	TGTGGGATTTCTGGAGCTTTCTGGGAAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGG	175	
DB	447	TGTGGGATTTCTGGAGCTTTCTGGGAAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGG	388	
QY	176	ACCCCTCGGGCTCAGAGCCCTAAAGTGGGCGCTGGTGAGCAGGCTGGTCTGGGCTCCAC	235	
DB	387	ACCCCTCGGGCTCAGAGCCCTAAAGTGGGCGCTGGTGAGCAGGCTGGTCTGGGCTCCAC	328	
QY	236	TTCCCAAGCCTCAGCAAGCTCATCTTCATTTGAATGCTCTCATTTGGCGGAGGACAACTG	295	
DB	327	TTCCCAAGCCTCAGCAAGCTCATCTTCATTTGAATGCTCTCATTTGGCGGAGGACAACTG	268	
QY	296	AACTTTGTGGTTGTGTTTACCCCTTCAAGTTGCTCCGCTCCCTACCCAGAGGTTTG	355	
DB	267	AACTTTGTGGTTGTGTTTACCCCTTCAAGTTGCTCCGCTCCCTACCCAGAGGTTTG	208	
QY	356	TCCGAGCCTGTGTCAGGGTTGTATAAACCAGGTAACCTTCGTTAGTTTGGCCATTC	415	
DB	207	TCCGAGCCTGTGTCAGGGTTGTATAAACCAGGTAACCTTCGTTAGTTTGGCCATTC	148	
QY	416	GCCATGGTCACGTGACATGCAAGTAATCTTGCTCTTAATTATAGAAATGATTTTCTTT	475	
DB	147	GCCATGGTCACGTGACATGCAAGTAATCTTGCTCTTAATTATAGAAATGATTTTCTTT	88	
QY	476	TAATTTTACTTTACCGACTTTACTTTGTTACTCAGAGAGAGCCCTCAGATGGCTGTG	535	
DB	87	TAATTTTACTTTACCGACTTTACTTTGTTACTCAGAGAGAGCCCTCAGATGGCTGTG	28	
QY	536	TCACATATAAATGTGGACTAAACTCT	562	
DB	27	TCACATATAAATGTGGACTAAACTCT	1	
RESULT 2				
AW136254/c				
LOCUS				
DEFINITION	455 bp mRNA linear EST 29-OCT-1999			
	UI-H-B11-act-e-12-0-UI-s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone			
	IMAGE:2715383 3', mRNA sequence.			
ACCESSION	AW136254			
VERSION	AW136254.1 GI:6140387			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1 (bases 1 to 455)			
	NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	The sequence contained an oligo-dT track that was present in the			
	oligonucleotide that was used to prime the synthesis of first			
	strand cDNA and therefore this may represent a bonafide poly A			
	tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:			
	NCI_CGAP clone distribution information can be found through the			
	I.M.A.G.E. Consortium/LLNL at:			
	www.bio.llnl.gov/bbrp/image/image.html			
	Seq primer: M13 Forward			
	POLYA=Yes.			
	Location/Qualifiers			
	1. .455			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:2715383"			
	/clone_lib="NCI_CGAP_Sub3"			
	/lab_host="DH10B (Life Technologies)"			
	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified			
	polylinker; Site_1: Not I; Site_2: Eco RI; The			
	NCI_CGAP_Sub3 library is a subtracted library derived from			
	the NCI_CGAP_Sub1 library, which is a subtracted library			
	derived from B1. B1 constitutes a mixture of 21			
	normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Col0,			
	NCI_CGAP_Pr22, NCI_CGAP_Kid5, NCI_CGAP_Kid12,			
	NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,			
	NCI_CGAP_Br2, NCI_CGAP_Col8, NCI_CGAP_Col1, NCI_CGAP_Le12,			
	NCI_CGAP_Br23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,			
	NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,			
	NCI_CGAP_Br25. These 21 libraries were pooled and a			
	single-stranded DNA preparation of the resulting mixture			
	was used as a tracer in a subtractive hybridization with			
	a driver whose composition is detailed below:			
	NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,			
	3798-3803 (IMAGE CloneIDs 132376-132391), 1456008-1456775			
	, 1500552-1502855; NCI_CGAP_Kid5 pool 1 LLAM 3338-3342			
	, 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,			
	1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1			
	LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,			
	1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,			
	3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,			
	1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1			
	LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs			
	985608-986759, 1101192-1101959, 1217928-1220615);			
	NCI_CGAP_Col0 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE			
	CloneIDs 1057416-1061255, 1144584-1145351). Subtraction			
	was performed as previously described [Bonaldo, Lennon &			
	Soares (1996): Normalization and Subtraction: Two			
	Approaches To Facilitate Gene Discovery. Genome Research			
	6, 791-806.			
	TAG_LIB=NCI_CGAP_Col0			
	TAG_TISSUE=colon			
	TAG_SEQ=AAACG			
BASE COUNT	137 a	102 c	101 g	115 t
ORIGIN				
	Query Match	77.8%;	Score 438;	DB 10; Length 455;
	Best Local Similarity	100.0%;	Pred. No. 6e-92;	
	Matches 438;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	126	TGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGATGCTCTCAAGGACCCCTCGG	185	
DB	455	TGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGATGCTCTCAAGGACCCCTCGG	396	
QY	186	CTCAGAGCCCTAAAGTGGCCCTGGTGAAGCAGGCTGCTCTCGTCCACTCCCAAGCC	245	

Db 395 CTCAGAGCCTAAAGTGGCCCTGGTGAAGCAGGCGTGTCTCGCTCCACTTCCCAAGCC 336
 QY 246 TGAGCCAGCTCATCTTCAATGATCTCAATTGGCCGAGGAACAACCTGAACCTTTGGG 305
 Db 335 TGAGCCAGCTCATCTTCAATGATCTCAATTGGCCGAGGAACAACCTGAACCTTTGGG 276
 QY 306 TTGTGCTTTAGCCTTCAGCTTTGCTCCGCTGCTCTACCCAGAGGTTTGGCGAGCCTG 365
 Db 275 TTGTGCTTTAGCCTTCAGCTTTGCTCCGCTGCTCTACCCAGAGGTTTGGCGAGCCTG 216
 QY 366 TGTTCAGGGTGTATATAACCAAGTACTTCGTTAGTTTTCGCCATTGAGCCATGTCFA 425
 Db 215 TGTTCAGGGTGTATATAACCAAGTACTTCGTTAGTTTTCGCCATTGAGCCATGTCFA 156
 QY 426 CGTGACATGCAAGTAACTTCTGCTCTAATATAGAATGATTTTCTTTAAATTTTAA 485
 Db 155 CGTGACATGCAAGTAACTTCTGCTCTAATATAGAATGATTTTCTTTAAATTTTAA 96
 QY 486 CTTTACAGACTTTACTTTGTACTCAGAGAAGGCGCTCAGATGGCTGTGCACATATAA 545
 Db 95 CTTTACAGACTTTACTTTGTACTCAGAGAAGGCGCTCAGATGGCTGTGCACATATAA 36
 QY 546 ATGTTGACTAAACTCTT 563
 Db 35 ATGTTGACTAAACTCTT 18

RESULT 3
 CNS0081M/C
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR16L12 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL051209
 VERSION AL051209.1 GI:4933162
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 896)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's F1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..896
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR16L12"
 /clone_lib="RPCI-98"
 /note="end : TET3"
 BASE COUNT 309 a 104 c 44 g 135 t 304 others
 ORIGIN

Query Match 7.0%; Score 39.6; DB 17; Length 896;

Best Local Similarity 26.3%; Pred. No. 42;
 Matches 64; Conservative 75; Mismatches 104; Indels 0; Gaps 0;
 QY 286 GGACAACTGAACCTTTGCTGTTGCTTTAGCCTTCAGCTTGCCTGCTCCTACC 345
 Db 860 GGATRCCTRSKMGVGBKRWTCGCGCKCGCMCMYCMYMMMSDSCTTDBASA 801
 QY 346 CAGAGGTTTGTGAGAGCCTGTGTTGCAGGTTGTATATAACCAAGGTTTCTGTTAGTT 405
 Db 800 CWSAMDTCSCGWACRAMAMWSAGAHAKAYATATADAAGMRVAVNGSGTKMKMKATTY 741
 QY 406 TGCCATTGACCATGCTCAGTGACATGCAAGTAACTTCTGCTCTAATTTATAGAAATG 465
 Db 740 MCSTBTYYTATWHTTSMYDITMWTATATSCRYATTTKTKTCTCTGCMGTMBASAMCK 681
 QY 466 ATTTTCTTTTAAATTTTACTTTTACAGACTTTTACTTTGTACTCAGAGAAGGCGCTCA 525
 Db 680 MGKGTGKTASKWCCTCWATMMRCASATATHIVACYHCTGTTCTVRMGAMACCCBKIDM 621
 QY 526 CAT 528
 Db 620 VVT 618

RESULT 4
 CNS021SK
 LOCUS
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 226G02 of library G from Tetraodon nigroviridis, genomic survey sequence.
 ACCESSION AL177293
 VERSION AL177293.1 GI:7815350
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 742)
 AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 742)
 AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 742)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000)
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source
 1..742
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="226G02"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG226BD01SPL-end : PUC-Ori"

BASE COUNT 140 a 114 c 121 g 226 t 141 others
 ORIGIN

Query Match 6.9%; Score 38.8; DB 17; Length 742;

```

Best Local Similarity 27.58; Pred. No. 65;
Matches 61; Conservative 63; Mismatches 98; Indels 0; Gaps 0;

QY 285 AGGAACAACTGAACATTTGTGGTTGCTGTTTACGCTTCAGTTTGCCTCGGTGCTCTAC 344
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 ATSMSSMGKSGSKAKSTGTTGTMGMWMTGSMGCTSKSTKAKTAKVTAKMKKKKSRASM 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 CCAGAGGTTTGGCAGCCTGTGTCAGAGGTTCTATAAACCACAGTACTTGGTAGTT 404
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 TTTMTTMTTSGGTRKAGSTTTTCMSTSKMTTWTMTMAWKGTTTTSMKTTSTT 267
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 405 TTGCCCATTCAGCCATGTCAGTCAGATGCAAGTAAATCTTGCTCTCCTAATATAGAAT 464
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 TWCMSMTTACMMCKMTTTTCTTGMKAAXSMWMSATSKTKTAAVAAKKACASCAT 327
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 465 GATTTTCTCTTAATTTTACTTACCAGACTTTTCTTGT 506
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 STTTTMTATMTSMBBSGSKTKSTATATTTGSSATGTT 369
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
A2255953/c
LOCUS A2255953 730 bp DNA linear GSS 26-JUL-2000
DEFINITION RPCI-23-111P7.TVB RPCI-23 Mus musculus genomic clone RPCI-23-111P7,
DNA sequence.
ACCESSION A2255953
VERSION A2255953.1 GI:9459216
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 730)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-111P7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 111 row: P column: 7
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..730
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-111P7"
/clone_lib="RPCI-23"
/sex="female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 290 a 95 c 114 g 231 t
ORIGIN

```

```

Query Match 6.88; Score 38.2; DB 17; Length 730;
Best Local Similarity 63.78; Pred. No. 90; Indels 0; Gaps 0;
Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 399 TTAGTTTGGCCATTCAGCCATGTCAGTCAGATGCAAGTAAATCTTGGCTCTAATAT 458
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 TTACTTTGGAGTATCAATATAGTCAAGTTTATGAAAAAATTTTCTCCTAATCAT 313
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 AGAAATGATTTTCTTTAAATTTTACTTT 489
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 AGAATTAATCCTTTGTTCAATGATGTTCAAT 282
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
A2986887/c
LOCUS A2986887 737 bp DNA linear GSS 27-APR-2001
DEFINITION 2M0269G13F Mouse 10kb plasmid U0GC2M library Mus musculus genomic
clone U0GC2M0269G13 F, DNA sequence.
ACCESSION A2986887
VERSION A2986887.1 GI:13858114
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 737)
Dunn,D., Aoyagi,A., Barber,M., Beaborn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0269 row: G column: 13
Seq primer: CGTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 737.
FEATURES
Location/Qualifiers
1..737
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GC2M0269G13"
/clone_lib="Mouse 10kb plasmid U0GC2M library"
/sex="female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (q14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

VERSION      BJ083396.1  GI:17578448
KEYWORDS     EST
SOURCE       African clawed frog.
ORGANISM     Xenopus laevis
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
              Xenopodinae; Xenopus.
REFERENCE    1 (bases 1 to 269)
AUTHORS      Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
              Y.
TITLE        Expressed genes in X. laevis embryo
JOURNAL      Unpublished (2001)
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.

FEATURES     source
              Location/Qualifiers
                1..269
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone="XL085a21"
                /clone_lib="NIBB Mochii normalized Xenopus tailbud
                library"
                /tissue_type="whole embryo"
                /dev_stage="stage 25"
                /note="vector: pBSRN3; Site1: NotI; Site2: EcoRI; cDNAs
                were oligo-dT primed and directionally cloned. Staging
                according to Nieukoop and Faber. Library is subtracted
                and was constructed by N. Garrett and A.M. Korn,
                (Wellcome/CRC Institute)."
              74 a 40 c 32 g 116 t 7 others
              BASE COUNT 74 a 40 c 32 g 116 t
              ORIGIN
                6.7% Score 37.8; DB 13; Length 269;
                Best Local Similarity 57.0%; Pred. No. 1.1e+02;
                Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
                QY 390 GGTACTTCGTAGTTTCCCATTCAGCGATGGTCAGCATGCAAGTAAATCTTGCT 449
                DB 12 GTTCTGTTGTTTAAACAGTCTTGGCCCTCCACAGGTCAGGAATAGAAAT 71
                QY 450 CCAATATAGAAATGATTTTCCTTTAAATTTTACCTTACCAGACTTTACTTTGACT 509
                DB 72 AATAATTAATAAACTCTCCCTCCCAAAATCTTCTTTAAATAACCTTATTTTCT 131
                QY 510 C 510
                DB 132 C 132
                RESULT 10
                BH106708 538 bp DNA linear GSS 19-JUL-2001
                LOCUS RPCI-24-401A14.TV RPCI-24 Mus musculus genomic clone RPCI-24-401A14
                DEFINITION DNA sequence.
                ACCESSION BH106708
                VERSION BH106708.1 GI:14937003
                KEYWORDS GSS.
                SOURCE house mouse.
                ORGANISM Mus musculus.
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
                1 (bases 1 to 538)
                REFERENCE Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
                Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
                Russell,D., de Jong,P. and Fraser,C.M.
                Mouse BAC End Sequences from Library RPCI-24
                Unpublished (1999)
                CONTACT: Shaying Zhao
                Department of Eukaryotic Genomics

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
Page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 401 row: A column: 14
Seq primer: T7
Class: BAC ends.

FEATURES     source
              Location/Qualifiers
                1..538
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPCI-24-401A14"
                /clone_lib="RPCI-24"
                /sex="Male"
                /cell_type="Spleen/Brain"
                /note="Vector: pTARBAC1; Site1: BamHI; Site2: BamHI;
                RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                library was cloned in the pTARBAC1 cloning vector at the
                BamHI sites using MboI partially digested male C57BL/6J
                DNA."
              202 a 83 c 85 g 168 t
              BASE COUNT 202 a 83 c 85 g 168 t
              ORIGIN
                6.7% Score 37.6; DB 17; Length 538;
                Best Local Similarity 62.3%; Pred. No. 1.2e+02;
                Matches 76; Conservative 0; Mismatches 44; Indels 2; Gaps 1;
                QY 427 GTGACATGCAAGTAATCTGCTCTAATTATAGAAA--TGATTTTCTTTTAAATTTT 484
                DB 395 GTGAATACAGAGCTCACTCACCCCTGATGTAATAAACTGATTTTGTG 454
                QY 485 ACTTTACAGACTTTTACTTTTACTTCAGAGAGAGGCTCACATGCGTGTGCATATA 544
                DB 455 CTGTTTATGCTTTATTTATTTATTAAGAAAACAAGCTGTCTATATCAGCAATAATAA 514
                QY 545 AA 546
                DB 515 AA 516
                RESULT 11
                BH56489 850 bp DNA linear GSS 19-FEB-2002
                LOCUS BOMJT12TF BO_2_3_KB Brassica oleracea genomic clone BOMJT12, DNA
                DEFINITION sequence.
                ACCESSION BH56489
                VERSION BH56489.1 GI:18714817
                KEYWORDS GSS.
                SOURCE Brassica oleracea.
                ORGANISM Brassica oleracea
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
                1 (bases 1 to 850)
                REFERENCE Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
                Whole genome shotgun sequencing of Brassica oleracea
                Unpublished (2001)
                CONTACT: Chris Town
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TF
                Class: sheared ends.

```

FEATURES	source	Location/Qualifiers
BASE COUNT	345 a	98 c 103 g 304 t
ORIGIN		
Query Match	6.6%;	Score 37.4; DB 17; Length 850;
Best Local Similarity	53.8%;	Pred. No. 1.4e-02;
Matches	77; Conservative	0; Mismatches 66; Indels 0; Gaps 0;
QY	376	TTGTATAAACCAAGTACTTCGTTAGTTTGGCCATTGCGCATGTCACATGTC 435
DB	780	TAAATATGTTATATGTTTCTGATTATCATTTTATAGGAATGGTTACTTGAATCA 721
QY	436	AAAGTATCTTCCTCAATATAGAAATGATTTCTTTTAAATTTTACTTTACCAGA 495
DB	720	AAATATTTTACTATATATAAATATATATATGTTATATATATTTCTTTTAAATGA 661
QY	496	CTTACTTTTCTACTACAGAGA 518
DB	660	TTACCATGTGAAGTCTCCCAACA 638
RESULT 12		
CNS01576		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
BASE COUNT	212 a	179 c 142 g 416 t 144 others
ORIGIN		
Query Match	6.6%;	Score 37.2; DB 17; Length 1093;
Best Local Similarity	34.1%;	Pred. No. 1.5e-02;
Matches	90; Conservative	43; Mismatches 131; Indels 0; Gaps 0;
QY	287	GAACAACGAACTTTGGTTTCTCTTTAGCGTTTCAGTTTCTCGCTGCTCTCTACCC 346

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:37:18 ; Search time 15.5464 Seconds
(without alignments)
11106.080 Million cell updates/sec

Title: US-09-914-152-3_COPY_10000_10562
Perfect score: 563
Sequence: 1 attcaccaccccaactgc.....aastgttggaactacttt 563

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	6.7	7218	1	US-08-232-463-14
2	32.8	5.8	40000	4	US-09-780-049-18
3	32.8	5.8	162450	4	US-09-345-882-1
4	32.6	5.8	960	3	US-09-130-749-1
5	32.6	5.8	960	3	US-09-130-749-1
6	32.4	5.8	58516	2	US-08-998-306-1
7	32.4	5.8	58516	4	US-09-338-907-1
8	32.4	5.8	58516	4	US-09-218-207-1
9	32.4	5.8	58520	4	US-09-338-907-179
10	32.4	5.8	58520	4	US-09-218-207-179
11	32.2	5.7	1530	5	PCT-US96-05320A-899
12	32.2	5.7	4852	1	US-07-853-913-3
13	31.8	5.6	925	4	US-09-257-179-17
14	31.2	5.5	1974	1	US-08-413-135-3
15	31.2	5.5	1974	4	US-08-971-395-3
16	31.2	5.5	2216	1	US-08-046-583-12
17	31.2	5.5	2216	1	US-08-384-556A-3
18	31.2	5.5	2216	1	US-08-331-355A-24
19	31.2	5.5	2216	5	PCT-US94-12364-24
20	31.2	5.5	2216	5	PCT-US95-07753-3
21	31.2	5.5	2298	3	US-09-157-077-7
22	31.2	5.5	2327	3	US-09-157-077-1
23	31	5.5	2238	2	US-08-674-887A-7
24	31	5.5	2238	3	US-08-951-844-7
25	31	5.5	2238	4	US-09-412-347-7
26	30.8	5.5	3279	4	US-08-445-137B-1
27	30.4	5.4	396	2	US-08-630-822A-84

Query Match 6.7% ; Score 37.6 ; DB 1 ; Length 7218 ;

c	28	30.4	5.4	396	2	US-09-005-069-84	Sequence 84, Appl
c	29	30.4	5.4	396	4	US-09-171-156A-33	Sequence 33, Appl
c	30	30.4	5.4	3410	4	US-09-651-656-42	Sequence 42, Appl
c	31	30.4	5.4	3410	4	US-09-650-855-42	Sequence 42, Appl
c	32	30.4	5.4	7215	4	US-09-134-001C-627	Sequence 627, Appl
c	33	30.2	5.4	234	4	US-09-134-001C-1635	Sequence 1635, Ap
c	34	30.2	5.4	599	4	US-09-328-111-147	Sequence 147, App
c	35	30.2	5.4	22067	4	US-09-820-001-3	Sequence 3, Appli
c	36	30	5.3	4089	4	US-09-134-001C-2118	Sequence 2118, Ap
c	37	29.8	5.3	1199	4	US-08-630-915A-15	Sequence 15, Appl
c	38	29.8	5.3	4526	1	US-07-855-412B-4	Sequence 4, Appli
c	39	29.8	5.3	4526	2	US-08-308-887A-4	Sequence 4, Appli
c	40	29.8	5.3	4526	3	US-08-881-094-4	Sequence 4, Appli
c	41	29.8	5.3	40328	3	US-08-742-185-102	Sequence 102, App
c	42	29.6	5.3	1740	2	US-08-864-224-10	Sequence 10, Appl
c	43	29.6	5.3	2346	4	US-09-193-503B-4	Sequence 4, Appli
c	44	29.6	5.3	2346	4	US-09-193-503B-7	Sequence 7, Appli
c	45	29.6	5.3	2346	4	US-09-193-503B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14


```
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
```

```
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
Query Match 5.8%; Score 32.8; DB 4; Length 162450;
Best Local Similarity 58.0%; Pred. No. 9.5;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 419 ATGGTCACGTGACATGCAAGTAATCTTGCTCCTAATTAGAAATGATTTTCTTTTAA 478
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125998 ATATTTCCTAAAGTACTAAGGAAAATATTTCCAAGGTATATATATGACTTGTCTGATAA 125939
QY 479 TTTTTCCTTACCAGACTTTTACTTTGTACTCAGAGAAGA 518
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125938 TATTTCCTTGGATCCTTACATTCACATCACTCAACAAACA 125899

RESULT 4
US-09-130-749-1
: Sequence 1, Application US/09130749
: Patent No. 6031090
: GENERAL INFORMATION:
: APPLICANT: SHABON, USMAN
: ELSHOURBAGY, NABIL
: TITLE OF INVENTION: MOLECULAR CLONING OF A 7TM
: RECEPTOR (GPR31A)
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RATNER & PRESTIA
: STREET: P.O. Box 980
: CITY: Valley Forge
: STATE: PA
```

COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/130,749
FILING DATE: 07-Aug-1998
CLASSIFICATION: UNKNOWN
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-70513
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-130-749-1

Query Match 5.8%; Score 32.6; DB 3; Length 960;
Best Local Similarity 51.7%; Pred. No. 0.78;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 213 AAGCAGGGTGGTCCTCGCTCCACITCCCAAGCCTGAGCCAGCTCATCTTCATTGAATGT 272
Db 590 ATGCAGGATCATAGGCTCTCCAGAAAGACTCCGGAGCCTGAGAAACAGCCCAAGC 649
QY 273 CTCATTTGGCCGAGAACACTGAACCTTGTGTTTGTCTTTAGCCTTCAGTTTGCTCC 332
Db 650 TTCAGCGGGCCGACGACTGCTACCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 709
QY 333 GCTGCTCTCTACCCAGAGGTTTG 355
Db 710 CTTGCTCTCTGCGCCAGAGTCTG 732

RESULT 5
US-09-130-749-1
Sequence 1, Application US/09130749
Patent No. 6031344
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: ELSHOURBAGY, NABIL
TITLE OF INVENTION: MOLECULAR CLONING OF A 7TM
TITLE OF INVENTION: RECEPTOR (GPR31A)
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/130,749

FILING DATE: 07-AUG-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-70513
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-130-749-1
Query Match 5.8%; Score 32.6; DB 3; Length 960;
Best Local Similarity 51.7%; Pred. No. 0.78;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 213 AAGCAGGGTGGTCCTCGCTCCACITCCCAAGCCTGAGCCAGCTCATCTTCATTGAATGT 272
Db 590 ATGCAGGATCATAGGCTCTCCAGAAAGACTCCGGAGCCTGAGAAACAGCCCAAGC 649
QY 273 CTCATTTGGCCGAGAACACTGAACCTTGTGTTTGTCTTTAGCCTTCAGTTTGCTCC 332
Db 650 TTCAGCGGGCCGACGACTGCTACCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 709
QY 333 GCTGCTCTCTACCCAGAGGTTTG 355
Db 710 CTTGCTCTCTGCGCCAGAGTCTG 732
RESULT 6
US-08-996-306-1/C
Sequence 1, Application US/08996306
Patent No. 5945522
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Chumakov, Ilya
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: Prostate cancer gene
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: win95
SOFTWARE: word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,306
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET 018A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 56516 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: DOUBLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: GENOMIC DNA
;   ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens
;   FEATURE:
;     NAME/KEY: Promoter
;     LOCATION: 1629..1870
;   IDENTIFICATION METHOD: Proscan
;   FEATURE:
;     NAME/KEY: Potential ATG
;     LOCATION: 1998..2000
;   FEATURE:
;     NAME/KEY: Exon 1
;     LOCATION: 2001..2216
;   FEATURE:
;     NAME/KEY: ATG
;     LOCATION: 2031..2033
;   FEATURE:
;     NAME/KEY: TYR phos
;     LOCATION: 11694..14332
;   FEATURE:
;     NAME/KEY: SEQ ID42
;     LOCATION: 11930..11947
;   FEATURE:
;     NAME/KEY: SEQ ID24
;     LOCATION: 12057..12103
;   FEATURE:
;     NAME/KEY: SEQ ID51
;     LOCATION: compl(12339..12358)
;   FEATURE:
;     NAME/KEY: SEQ ID64
;     LOCATION: 13547..13564
;   FEATURE:
;     NAME/KEY: SEQ ID58
;     LOCATION: 13657..13703
;   FEATURE:
;     NAME/KEY: SEQ ID67
;     LOCATION: compl(13962..13981)
;   FEATURE:
;     NAME/KEY: Exon 2
;     LOCATION: 18196..18265
;   FEATURE:
;     NAME/KEY: Exon 3
;     LOCATION: 23717..23832
;   FEATURE:
;     NAME/KEY: Exon 4
;     LOCATION: 25571..25660
;   FEATURE:
;     NAME/KEY: SEQ ID43
;     LOCATION: 34216..34234
;   FEATURE:
;     NAME/KEY: SEQ ID25
;     LOCATION: 34469..34515
;   FEATURE:
;     NAME/KEY: SEQ ID52
;     LOCATION: compl(34625..34645)
;   FEATURE:
;     NAME/KEY: Exon 5
;     LOCATION: 34669..34759
;   FEATURE:
;     NAME/KEY: Exon 6
;     LOCATION: 40688..40846
;   FEATURE:
;     NAME/KEY: Exon 7
;     LOCATION: 48070..48193
;   FEATURE:
;     NAME/KEY: Exon 8
;     LOCATION: 50182..54523

```

```

; FEATURE:
;   NAME/KEY: SEQ ID65
;   LOCATION: 51149..51168
;   FEATURE:
;   NAME/KEY: SEQ ID59
;   LOCATION: 51448..51494
;   FEATURE:
;   NAME/KEY: SEQ ID68
;   LOCATION: compl(51482..51499)
;   FEATURE:
;   NAME/KEY: SEQ ID44
;   LOCATION: 51596..51613
;   FEATURE:
;   NAME/KEY: SEQ ID26
;   LOCATION: 51612..51658
;   FEATURE:
;   NAME/KEY: SEQ ID53
;   LOCATION: compl(51996..52015)
;   FEATURE:
;   NAME/KEY: polyad signal
;   LOCATION: 54445..54450
;   US-08-996-306-1
;
; Query Match          5.8%; Score 32.4; DB 2; Length 56516;
; Best Local Similarity 50.0%; Pred. No. 7.4;
; Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
;
; QY 380 ATAAACCAAGGACTTCGTTAGTTTGGCCCATTCAGCCATGGTCACGTGACATGCAAG 439
;      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
; Db 8604 ACAATGTCACCTGCTCAGTAAGATTTCCAAACCCCTACCTCTCACCCACGAGCTAA 8545
;
; QY 440 TAATCTTGCTCCTAATTATAGAAATGATTTTCTTTTAAATTTTACTTTTACCAGACTTT 499
;      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
; Db 8544 CRACCTGGCATGGCCAGCAGTAATTTTCATTTTCTTGTTATTTGTTGTTCTTCCACACTAG 8485
;
; QY 500 ACTTGTACTCAGAGAGAGGCGCTCACATGGCTGTGTGCACAT 541
;      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
; Db 8484 AATGTACCTCCATAATGTCAGGTACATTTGTCTATCAACAT 8443
;
; RESULT 7
; US-09-338-907-1/c
;   Sequence 1, Application US/09338907
;   Patent No. 6265546
;   GENERAL INFORMATION:
;     APPLICANT: Cohen, Daniel
;     APPLICANT: Blumenfeld, Marta
;     APPLICANT: Iliya, Chumakov
;     APPLICANT: Bougueleret, Lydie
;     TITLE OF INVENTION: PROSTATE CANCER GPNE
;     FILE REFERENCE: GENSET.18CP1CP
;     CURRENT APPLICATION NUMBER: US/09/338.907
;     CURRENT FILING DATE: 1999-06-23
;     EARLIER APPLICATION NUMBER: 08/996.306
;     EARLIER FILING DATE: 1997-12-22
;     EARLIER APPLICATION NUMBER: 60/099.658
;     EARLIER FILING DATE: 1998-09-09
;     EARLIER APPLICATION NUMBER: 09/218.207
;     EARLIER FILING DATE: 1998-12-22
;     NUMBER OF SEQ ID NOS: 578
;     SOFTWARE: Patent.pm
;     SEQ ID NO 1
;     LENGTH: 56516
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
;     NAME/KEY: promoter
;     LOCATION: 1629..1870
;   OTHER INFORMATION: Identification method Proscan
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 1998..2000
;   OTHER INFORMATION: potential start codon

```

```
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2031..2033
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11694..114332
OTHER INFORMATION: Tyr Phos
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
FEATURE:
NAME/KEY: allele
LOCATION: 12057..12103
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
FEATURE:
NAME/KEY: allele
LOCATION: 13657..13703
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon 2
FEATURE:
NAME/KEY: exon
LOCATION: 23717..23832
OTHER INFORMATION: exon 3
FEATURE:
NAME/KEY: exon
LOCATION: 25571..25660
OTHER INFORMATION: exon 4
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34216..34234
OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
FEATURE:
NAME/KEY: allele
LOCATION: 34469..34515
OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34625..34645
OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
FEATURE:
NAME/KEY: exon
LOCATION: 34669..34759
OTHER INFORMATION: exon 5
FEATURE:
NAME/KEY: exon
LOCATION: 40888..40846
OTHER INFORMATION: exon 6
FEATURE:
NAME/KEY: exon
LOCATION: 48070..48193
OTHER INFORMATION: exon 7
FEATURE:
```

```
NAME/KEY: exon
LOCATION: 50182..54523
OTHER INFORMATION: exon 8
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51149..51168
OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: allele
LOCATION: 51448..51494
OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51482..51499
OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, compleme
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51596..51613
OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
FEATURE:
NAME/KEY: allele
LOCATION: 51612..51658
OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51996..52015
OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, compleme
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 54445..54450
OTHER INFORMATION: AATAAA
US-09-338-907-1

Query Match      5.8%; Score 32.4; DB 4; Length 56516;
Best Local Similarity 50.0%; Pred. No. 7.4;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 380 ATAAACCAAGGTACTTCGGTTAGTTTGGCCCATTCACGCCATGGTCAGTCACATGCAAG 439
DB 8604 ACAATGTCACTTGTCTCAGTAAGATTTTCCAAACCCCTACCTCTCACCACCAAGCTAA 8545
QY 440 TAATCTTGCTCCTAATTATAGAAATGATTTCTTTTAAATTTTACTTTTACCAGACTTT 499
DB 8544 CAACCTGGCATGGCAGACAGTAAATTTTCATTTTCTTGTATTGTTGTCCTCCACACTAG 8485
QY 500 ACTTTGTACTCAGAGAAGGCCCTCACATGGCTGTGTACAT 541
DB 8484 AATGTGACCTCCATAATATGGCAGGTACATTTGCTATCAACAT 8443

RESULT 8
US-09-218-207-1/c
: Sequence 1, Application US/09218207
: Patent No. 6346381
: GENERAL INFORMATION:
: APPLICANT: Cohen, Daniel
: APPLICANT: Blumenfeld, Marta
: APPLICANT: Ilya, Chumakov
: APPLICANT: Bougueleret, Lydie
: TITLE OF INVENTION: Prostate cancer gene
: FILE REFERENCE: GENSET.018Gp1
: CURRENT APPLICATION NUMBER: US/09/218,207
: EARLIER FILING DATE: 1998-12-22
: EARLIER APPLICATION NUMBER: 08/996,306
: EARLIER FILING DATE: 1997-12-22
: EARLIER APPLICATION NUMBER: 60/099,658
: EARLIER FILING DATE: 1998-09-09
: NUMBER OF SEQ ID NOS: 578
: SOFTWARE: Patent.pm
: SEQ ID NO 1
: LENGTH: 56516
: TYPE: DNA
: ORGANISM: Homo sapiens
```

FEATURE:
NAME/KEY: promoter
LOCATION: 1629..1870
OTHER INFORMATION: identification method Proscan
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2000
OTHER INFORMATION: potential start codon
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2031..2033
OTHER INFORMATION: AUG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11694..14332
OTHER INFORMATION: Tyr Phos
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
FEATURE:
NAME/KEY: allele
LOCATION: 12057..12103
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
FEATURE:
NAME/KEY: allele
LOCATION: 13657..13703
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon 2
FEATURE:
NAME/KEY: exon
LOCATION: 23717..23832
OTHER INFORMATION: exon 3
FEATURE:
NAME/KEY: exon
LOCATION: 25571..25660
OTHER INFORMATION: exon 4
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34216..34234
OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
FEATURE:
NAME/KEY: allele
LOCATION: 34469..34515
OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34625..34645
OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
FEATURE:
NAME/KEY: exon
LOCATION: 34669..34759
OTHER INFORMATION: exon 5
FEATURE:

NAME/KEY: exon
LOCATION: 40688..40846
OTHER INFORMATION: exon 6
FEATURE:
NAME/KEY: exon
LOCATION: 48070..48193
OTHER INFORMATION: exon 7
FEATURE:
NAME/KEY: exon
LOCATION: 50182..54523
OTHER INFORMATION: exon 8
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51149..51168
OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: allele
LOCATION: 51448..51494
OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51482..51499
OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51596..51613
OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
FEATURE:
NAME/KEY: allele
LOCATION: 51612..51658
OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51996..52015
OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
FEATURE:
NAME/KEY: polyA-signal
LOCATION: 54445..54450
OTHER INFORMATION: AATAAA
US-09-218-207-1
Query Match 5.8%; Score 32.4; DB 4; Length 56516;
Best Local Similarity 50.0%; Pred. No. 7.4;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 380 ATAAACCAAGGTACTTCGTTAGTTTTCGCCATTGCGCATGTCAGTGCATGCAAG 439
Db 8604 ACAATGTCACCTGCTCAGTAGATTTTCCAAACCCCTACCTCTCACCACCGCTAAA 8545
QY 440 TAATCTTGCTCTAATTATAGAAATGATTTTCTTTTAAATTTTACTTTTACCAGATTT 499
Db 8544 CAACCTGGCATGGCACAGTAAATTTTCATTTTCTTTGTTATTTGTTCTTCCACACTAG 8485
QY 500 ACTTTGTACTCAGACAGAGCCCTCACATGGCTGTGTGCACAT 541
Db 8484 AATGTGACCTCCATAATGCGAGGTACATTGTCTATCAAAACAT 8443
RESULT 9
US-09-338-907-179/c
Sequence 179, Application US/09338907
Patent No. 6265346
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CPICP
CURRENT APPLICATION NUMBER: US/09/338,907
CURRENT FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22


```

; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 09/218,207
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 179
; LENGTH: 56520
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
;
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon2
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23716..23831
; OTHER INFORMATION: exon3
;
; NAME/KEY: exon
; LOCATION: 25570..25659
; OTHER INFORMATION: exon4
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34668..34758
; OTHER INFORMATION: exon5
;
; NAME/KEY: exon
; LOCATION: 40685..40843
; OTHER INFORMATION: exon6
;
; NAME/KEY: exon
; LOCATION: 48067..48190
; OTHER INFORMATION: exon7
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50179..54519
; OTHER INFORMATION: exon8
;
; NAME/KEY: polyA_signal
; LOCATION: 54493..54498
; OTHER INFORMATION: AATAAA
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1991..2008
; OTHER INFORMATION: upstream amplification primer 5-63
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2505..2525
; OTHER INFORMATION: downstream amplification primer 5-63 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4091..4111
; OTHER INFORMATION: downstream amplification primer 99-622
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4528..4546
; OTHER INFORMATION: upstream amplification primer 99-622 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5475..5495
; OTHER INFORMATION: downstream amplification primer 99-621
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5927..5947
; OTHER INFORMATION: upstream amplification primer 99-621 , complement
;
; NAME/KEY: primer_bind
; LOCATION: 8127..8144
; OTHER INFORMATION: downstream amplification primer 99-607
;
;
; OTHER INFORMATION: downstream amplification primer 99-619
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8560..8578
; OTHER INFORMATION: upstream amplification primer 99-619 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11622..11639
; OTHER INFORMATION: upstream amplification primer 4-76
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12018..12037
; OTHER INFORMATION: downstream amplification primer 4-76 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12915..12932
; OTHER INFORMATION: upstream amplification primer 4-71
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13317..13334
; OTHER INFORMATION: downstream amplification primer 4-71 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13216..13233
; OTHER INFORMATION: upstream amplification primer 4-72
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13617..13636
; OTHER INFORMATION: downstream amplification primer 4-72 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13547..13564
; OTHER INFORMATION: upstream amplification primer 4-73
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13962..13981
; OTHER INFORMATION: downstream amplification primer 4-73 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 15994..16011
; OTHER INFORMATION: downstream amplification primer 99-610
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 16463..16480
; OTHER INFORMATION: upstream amplification primer 99-610 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 17304..17324
; OTHER INFORMATION: downstream amplification primer 99-609
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 17814..17832
; OTHER INFORMATION: upstream amplification primer 99-609 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18008..18027
; OTHER INFORMATION: upstream amplification primer 4-90
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18423..18442
; OTHER INFORMATION: downstream amplification primer 4-90 , complement
;
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 18699..18716
; OTHER INFORMATION: downstream amplification primer 99-607
;

```



```
FEATURE:
NAME/KEY: primer_bind
LOCATION: 8560..8578
OTHER INFORMATION: upstream amplification primer 99-619, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11622..11639
OTHER INFORMATION: upstream amplification primer 4-76
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12018..12037
OTHER INFORMATION: downstream amplification primer 4-76, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12915..12932
OTHER INFORMATION: upstream amplification primer 4-71
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13317..13334
OTHER INFORMATION: downstream amplification primer 4-71, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13216..13233
OTHER INFORMATION: upstream amplification primer 4-72
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13617..13636
OTHER INFORMATION: downstream amplification primer 4-72, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
```

```
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
OTHER INFORMATION: upstream amplification primer 99-592, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34624..34644
```

Query Match 5.8%; Score 32.4; DB 4; Length 56520;
Best Local Similarity 50.0%; Pred. No. 7.4;
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
QY 380 ATAAACCAAGGTACTTCGTTAGTTTTGCCCATTCAGCCATGGTCAGTGACATGCAAG 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8604 ACAATGTCACTTCTCAGTAAGATTTCCAAACCCCTACCTCACCACCGCTAAA 8545
QY 440 TAATCTTCTCCTAATTATAGAATGATTTTCTTTTAAATTTTACTTTTACCAGACTT 499
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8544 CAACCTGGCATGGCACAGTAAATTTTCATTTCTTGTATTTGTTCTTCCACACTAG 8485
QY 500 ACTTTGTACTCAGAGAAGAGCGCTCACATGGCTGTGCACAT 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8484 AATGTGACCTCCATAATGGCAGGTACATTTGCTATCAACAT 8443
```

RESULT 11

PCT-US96-05320A-899/c
; Sequence 899, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams

```
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Disette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 899:
SEQUENCE CHARACTERISTICS:
LENGTH: 1530 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-899

Query Match 5.7%; Score 32.2; DB 5; Length 1530;
Best Local Similarity 59.1%; Pred. No. 1.3;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 234 ACTTCCAGCGCTGAGCAAGCTCATCTTCATTGAATGCTCTCATTTGGCGGAGGAACAAC 293
Db 657 ACATCCAGCGCTAAGACAAGTAAACTAAACCGACTTATCAATAGTTGATGAACAAT 598

QY 294 TGAACCTTTGTTGCTGTTTAGCCTTCAGTT 326
Db 597 TTCACCTTCTGCTTCCCAAAATCTCCAGCT 565

RESULT 12
US-07-853-913-3
; Sequence 3, Application US/07853913
; Patent No. 5318839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-853-913-3

Query Match 5.7%; Score 32.2; DB 1; Length 4852;
Best Local Similarity 50.3%; Pred. No. 2.4;
Matches 79; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 106 ATGGGGTTCATGTGGAGTTCTGGAGCTTCTGGGAATTCAGTTGGAGTCAGGAT 165
Db 3095 AGGGAGGGGCTGAGGCGCTCCAGGACCTCAAGGCAATCACACAGTGGGGACCCAG 3154

QY 166 GCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGCCCTGGTGAAGCAGGTGTC 225
Db 3155 GCTCCAGGCTCCCGAGGGCTGCCAGAGCGGATAGAGCCCTGGTGGAGATGATGTGG 3214

QY 226 CTGCGTCCACTTCCCAAGCCTGAGCCAAAGCTCATCTT 262
Db 3215 CCCAGGGGGTGACCAAGCCTCCCGAGAGTCACTTT 3251

RESULT 13
US-09-257-179-17/c
; Sequence 17, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: PZ015P1
; CURRENT APPLICATION NUMBER: US/09/257,179
; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 925
TYPE: DNA
ORGANISM: Homo sapiens
US-09-257-179-17

Query Match 5.5%; Score 31.8; DB 4; Length 925;
Best Local Similarity 64.0%; Pred. No. 1.4; 27; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 0; Mismatches 0; Mismatches 0;

QY 294 TGAACCTTGTGGTGTAGCTTTCAGTTTCCTCCGCTGCTCTACCCAGAGTTT 353

DB 239 TGTCTTGGAGTCCCTCTTGGCTTCCCTTGGAGTCCGCGCATCTTTCCTAGAGAT 180

QY 354 TGTGGAGCCTGTGT 368

DB 179 GCTAGGAGGGTGTGT 165

RESULT 14
US-08-413-135-3/c
Sequence 3, Application US/08413135
Patent No. 5689042

GENERAL INFORMATION:
APPLICANT: Amasino, Richard M
APPLICANT: Gan, Sushang
TITLE OF INVENTION: Transgenic Plants with Altered
TITLE OF INVENTION: Senescence Characteristics
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI

COUNTRY: US

ZIP: 53703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/413,135

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27,386

REFERENCE/DOCKET NUMBER: 960296.92808

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1974 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "SAG13 Promoter DNA"

US-08-413-135-3

Query Match 5.5%; Score 31.2; DB 1; Length 1974;
Best Local Similarity 58.7%; Pred. No. 3.2;
Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 411 ATTCAGCATGTCACGTGACATGCAAGTAATCTCTCTTAATTATAGAAATGATTT 470

DB 1289 ATTCTTCCATCTCTTTTAAATGCAATGTAATCTCTCTTTCTTTCTTTTCTTTT 1230

QY 471 TCTTTTAAATTTTACTTTTACCAGACITTTACT 502

DB 1229 TTTTCTTTTCTTAAAGATTGCTCT 1198

Db 1229 TTTTCTTTTCTTAAAGATTGCTCT 1198

RESULT 15

US-08-971-395-3/c

Sequence 3, Application US/08971395

Patent No. 6359197

GENERAL INFORMATION:

APPLICANT: Amasino, Richard M

APPLICANT: No. 6359197, Yoo-Sun

APPLICANT: Gan, Sushang

TITLE OF INVENTION: Transgenic plants with Altered

TITLE OF INVENTION: Senescence Characteristics

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/971,395

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.94908

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1974 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-971-395-3

Query Match 5.5%; Score 31.2; DB 4; Length 1974;

Best Local Similarity 58.7%; Pred. No. 3.2;

Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 411 ATTCAGCATGTCACGTGACATGCAAGTAATCTCTCTTAATTATAGAAATGATTT 470

DB 1289 ATTCTTCCATCTCTTTTAAATGCAATGTAATCTCTCTTTCTTTCTTTTCTTTT 1230

QY 471 TCTTTTAAATTTTACTTTTACCAGACITTTACT 502

DB 1229 TTTTCTTTTCTTAAAGATTGCTCT 1198

Search completed: April 11, 2003, 22:52:28

Job time : 252.546 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 21:30:39 ; Search time 36.0203 Seconds
(without alignments)
13710.183 Million cell updates/sec

Title: US-09-914-152-3_COPY_10000_10562
Perfect score: 563
Sequence: 1 attcatcacccaactgtc.....aaatgttggaactacttt 563

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	34.6	6.1	521	10	US-09-864-761-12861 Sequence 12861, A
C 2	34.4	6.1	564	10	US-09-864-761-13686 Sequence 13686, A
C 3	34.4	6.1	564	10	US-09-864-761-30198 Sequence 30198, A
C 4	34.2	6.1	12123	9	US-09-764-872-752 Sequence 752, App
C 5	33.2	5.9	15987	9	US-10-092-154-1301 Sequence 1301, App
C 6	33.2	5.9	15987	10	US-09-764-847-1301 Sequence 1301, App
C 7	33.2	5.9	172637	10	US-09-805-458A-3 Sequence 3, Appli
C 8	32.8	5.8	28197	9	US-10-092-154-1965 Sequence 1965, App
C 9	32.8	5.8	28197	10	US-09-764-847-1965 Sequence 1965, App
C 10	32.8	5.8	28210	9	US-10-092-154-1966 Sequence 1966, App
C 11	32.8	5.8	28210	10	US-09-764-847-1966 Sequence 1966, App
C 12	32.6	5.8	880	9	US-09-981-353-27 Sequence 27, Appl
C 13	32.6	5.8	2061	10	US-09-962-832-239 Sequence 239, App
C 14	32.6	5.8	2061	10	US-09-954-456-1593 Sequence 1593, App
C 15	32.4	5.8	862	10	US-09-974-300-2677 Sequence 2677, App
C 16	32.4	5.8	2000	9	US-09-938-842A-3003 Sequence 3003, App
C 17	32.4	5.8	56516	9	US-09-853-526-1 Sequence 1, Appli
C 18	32.4	5.8	56516	10	US-09-901-484A-1 Sequence 1, Appli
C 19	32.4	5.8	56520	9	US-09-853-526-179 Sequence 179, App

ALIGNMENTS

RESULT 1

US-09-864-761-12861/c

; Sequence 12861, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL

; FILE REFERENCE: Aemica-X-1

; CURRENT APPLICATION NUMBER: US 09/864,761

; PRIOR FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/006666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006661

; PRIOR FILING DATE: 2001-01-30

Sequence 179, App
Sequence 800, App
Sequence 2, Appli
Sequence 1, Appli
Sequence 7, Appli
Sequence 12, Appli
Sequence 17, Appli
Sequence 9, Appli
Sequence 41, Appli
Sequence 1298, Ap
Sequence 1298, Ap
Sequence 8512, Ap
Sequence 1, Appli
Sequence 4380, Ap
Sequence 306, App
Sequence 306, App
Sequence 6499, Ap
Sequence 40, Appli
Sequence 40, Appli
Sequence 177, App
Sequence 5032, Ap
Sequence 7, Appli
Sequence 127, App
Sequence 261, App
Sequence 261, App
Sequence 940, App

```

286 A U C C A T C A T T T T C A T G A C C T C G T G T C A C C T C C T G G T C C A C C C C C A T C T T C T T C A T 227
QY 444 C T G T C C T C A A T T A G A A A T G A T T T T C T T T A A T T T T T T A C T T T T A C C A G A C T T T A C T T 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 C T G G A T C A T T A T A A G G T T T C T A A G T G G T C T T C A T T T C T T C C C T T A T T C C C C T T A G T G 167
QY 504 T G T A C T C A G A G A G G C C T C A T G G T G T G T C A C A T A T A A T 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 T A C C C C C C C A G T A G G C A G A G A T C A T T T A C A T A T G A T 123

RESULT 3
US-09-864-761-30198/c
; Sequence 30198, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US/60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US/09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04

```

```
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30198
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049649.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: NT HIT: AF02236.1, EVALUE 9.00e-02
; OTHER INFORMATION: EST_HUMAN HIT: C03299.1, EVALUE 5.30e-01
US-09-864-761-30198

Query Match
Best Local Similarity 6.1%; Score 34.4; DB 10; Length 564;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 384 AACCAAGTACTTCCTGTTAGTTTCCCATTCAGCCATGGTCACGTGACATGCAAGTAAT 443
DB 286 ATCCATCATTTTTCATGACCTCTGTCACCTCCCTGGTCCACCCCATCTTCTCAT 227

QY 444 CTGCTCTTAATTAGAAATGATTTTCTTTTAAATTTTACTTTTACCAGACTTACTT 503
DB 226 CPGATCATTAATAGATTTCTTAAGTGGTCTTCATTTCTCCCTTATCCCTTTAGTG 167

QY 504 TGCTACTCAGAGAAGGCGCTCACATGGCTGTGCACATATAAAT 547
DB 166 TACCCCCAGCAAGTAGGCACAGAGATCCATTTAACAATACGT 123

RESULT 4
US-09-764-872-752
; Sequence 752, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
```

```
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 752
; LENGTH: 12123
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-752

Query Match
Best Local Similarity 6.1%; Score 34.2; DB 9; Length 12123;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 124 TCTGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGAGACCCCTCG 183
DB 10413 TCTGACCTCTCCAGTGAGGAGGCTCAGGGAAGGAGGATGCTCTGCTGGAGCCTGC 10472

QY 184 GCGTCAGAGCCCTAAAGTGGGCCCTGGTGAACCAAGGTTGCTCCGCTCCA 234
DB 10473 AGCCCCAGCCCTGCAGTGGGAGGAGGTCTCTGGGGTCTGCTTCTCCCA 10523

RESULT 5
US-10-092-154-1301/c
; Sequence 1301, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1301
; LENGTH: 15987
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1301

Query Match
Best Local Similarity 5.9%; Score 33.2; DB 9; Length 15987;
Matches 44; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 214 ACAGGGGTGCTCTCGTCCACTTCCCAAGCTGAGCCAGCTCATCTTCTTGAATGTC 273
DB 13116 ACAGGGAGGTCCCGCCAGCTCCCATCCCAAGAACCCAGCTACCTTACCTTGAGAGGC 13057

QY 274 TC 275
DB 13056 TC 13055

RESULT 6
US-09-764-847-1301/c
; Sequence 1301, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1301
; LENGTH: 15987
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1301
```



```

Qy 434 GCAAGTAATCTTGCTCTAATATTAGAATGATTTTCTTTTAATTTTTTACTTT 489
|| |||| || |||| || |||| || |||| || |||| || |||| || |||| ||
Db 11027 GCCTTTAAATTTAATCAGAAATTTTTCGAATTAATTTAAAAAATATTTTAACTT 11082

RESULT 11
US-09-764-847-1966
; Sequence 1966, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1966
; LENGTH: 26210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1966

Query Match 5.8%; Score 32.8; DB 10; Length 26210;
Best Local Similarity 55.2%; Pred. No. 43;
Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 374 GGTGTGTTAAACCAAGTACTTCGTTAGTITTTGCCCATTCAGCCATGTCACGTGACAT 433
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 10967 GGTTCFAGAAATTTAAATTTCTTAGTGATTTTAAAGCACTTACATGTTCCCACTGAAGTG 11026

Qy 434 GCAAGTAATCTTGCTCTAATATTAGAATGATTTTCTTTTAATTTTTTACTTT 489
|| |||| || |||| || |||| || |||| || |||| || |||| || |||| ||
Db 11027 GCCTTTAAATTTAATCAGAAATTTTTCGAATTAATTTAAAAAATATTTTAACTT 11082

RESULT 12
US-09-981-353-27
; Sequence 27, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 27
; LENGTH: 880
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 998971.1
US-09-981-353-27

Query Match 5.8%; Score 32.6; DB 9; Length 880;
Best Local Similarity 55.9%; Pred. No. 8.4;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 436 AAGTAATCTTGCTCTAATATTAGAATGATTTTCTTTTAATTTTTTACTTTTACCAGA 495
|||| |||| || |||| || |||| || |||| || |||| || |||| || |||| ||
Db 740 AATCAATTCAGAAATCTGTTAAACACACTATTATTATTTTATTTTAAATTAATATA 799

Qy 496 CTTTACTTTGTACTCAGAGAAGCGCTCACATGGCTGTGCACATATAAA 546
|| |||| || |||| || |||| || |||| || |||| || |||| || |||| ||
Db 800 ATATTCGTCTTATCTCTAAAGACAGATTAAAAATTTGTTGATAAATATAA 850

RESULT 13

```

```
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1593

Query Match      5.8%; Score 32.6; DB 10; Length 2061;
Best Local Similarity 51.7%; Pred. No. 13;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 213 AAGCAGGTTGCTCGTCCACTTCCCAAGCTGAGCCAGCTCATCTTCAATGATGT 272
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1088 ATGAGGATCATCAGGGCTCTCCAGAAAGACTCGGGAGGCTGAGAAACACCCCAAGC 1147
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 273 CTATTTGGCGGAGGAACTCAACTTTGTGTTTGTCTTTAGCTTCAGTTTGTCC 332
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1148 TTCAGCGGGCCAGGACTGGTACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1207
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 333 GCTGCTCTCTACCCAGAGTTTG 355
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1208 CTTGCTTCTGTGCCAGAGTCTG 1230

RESULT 15
US-09-974-300-2677/c
; Sequence 2677, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2677
; LENGTH: 862
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2677

Query Match      5.8%; Score 32.4; DB 10; Length 862;
Best Local Similarity 50.6%; Pred. No. 9.6;
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 326 TTGCTCCGCTGCTCTACCCAGAGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAA 385
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 286 TTGCGAGTATCTCTCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 386 CCAAGTACTTGTAGTTTGCCTTCCAGCATGTCAGTGTGACATGCAAGTAATCT 445
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 226 CATAGTTATAGCGGATTTTATCTACTCATCCAGAGCGAATCGGAATTTTGACATACTT 167
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 446 TGCTCCTAATATTAGNAATGATTTTCTTTAAT 479
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 166 TGCCGTTGGACATAATGATGCTTCTTCTGTAAT 133
```

Search completed: April 12, 2003, 04:36:37
Job time : 144.02 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:52:34 ; Search time 1356.77 Seconds
(without alignments)
16087.603 Million cell updates/sec

Title: US-09-914-152-3_COPY_1_750
Perfect score: 750
Sequence: 1 cgcctctgcaaggtagacc.....ctagaatagaattgttcgt 750

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vt:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	750	100.0	10562	6	E38420	E38420 Novel polyp
2	750	100.0	170121	9	AF064860	AF064860 Homo sapi
3	750	100.0	340000	9	HS21C080	AL163280 Homo sapi
4	279	37.2	185982	2	AC073231	AC073231 Homo sapi
5	127	16.9	185982	2	AC073231	AC073231 Homo sapi
6	24	3.2	4368	6	E06819	E06819 DNA encodin
7	24	3.2	4368	6	I19108	I19108 Sequence 17
8	24	3.2	4398	6	E06594	E06594 cDNA encodi
9	24	3.2	4446	6	I19102	I19102 Sequence 6
10	24	3.2	4449	10	MUSGRP2	DI0651 Mouse mRNA
11	24	3.2	235126	2	AC124590	AC124590 Mus muscu
12	23	3.1	168407	2	AC090786	AC090786 Homo sapi
13	22	2.9	140356	9	AC007628	AC007628 Genomic s
14	22	2.9	141485	9	AC009758	AC009758 Homo sapi
15	22	2.9	147318	2	AC079759	AC079759 Homo sapi
16	22	2.9	148313	9	AC010930	AC010930 Homo sapi
17	22	2.9	186654	2	AP001772	AP001772 Homo sapi
18	22	2.9	196320	2	AC124389	AC124389 Mus muscu
19	22	2.9	200829	2	AC090208	AC090208 Homo sapi
20	22	2.9	217090	2	AC117238	AC117238 Mus muscu
21	21	2.8	831	6	AX121096	AX121096 Sequence
22	21	2.8	954	6	AX065869	AX065869 Sequence
23	21	2.8	954	6	AX066697	AX066697 Sequence
24	21	2.8	3736	9	HSAT2739	AJ002739 Homo sapi
25	21	2.8	6167	6	AX323756	AX323756 Sequence
26	21	2.8	34576	3	CET2308	Z81128 Caenorhabdi
27	21	2.8	62964	2	AC130137	AC130137 Rattus no
28	21	2.8	73575	10	AL645766	AL645766 Mouse DNA
29	21	2.8	88658	9	AC110087	AC110087 Homo sapi
30	21	2.8	89856	2	AC111985	AC111985 Rattus no
31	21	2.8	110000	2	AC122879_1	Continuation (2 of
32	21	2.8	116993	9	AC073336	AC073336 Homo sapi
33	21	2.8	138636	9	HS1045C21	AL021919 Human DNA
34	21	2.8	153823	9	AC016623	AC016623 Homo sapi
35	21	2.8	165354	9	AC113368	AC113368 Homo sapi
36	21	2.8	165597	2	AC109706	AC109706 Rattus no
37	21	2.8	167803	2	AC100798	AC100798 Homo sapi
38	21	2.8	168521	9	AC093260	AC093260 Homo sapi
39	21	2.8	169678	2	AF301237	AF301237 Homo sapi
40	21	2.8	169845	9	AC105082	AC105082 Homo sapi
41	21	2.8	173380	9	AC087348	AC087348 Homo sapi
42	21	2.8	174336	2	AC123929	AC123929 Mus muscu
43	21	2.8	193254	9	AC008945	AC008945 Homo sapi
44	21	2.8	196853	2	AC023126	AC023126 Homo sapi
45	21	2.8	198772	10	AL606906	AL606906 Mouse DNA

ALIGNMENTS

RESULT 1
E38420
LOCUS E38420 Novel polypeptide.
DEFINITION E38420 Novel polypeptide.
ACCESSION E38420
VERSION E38420.1 GI:18626994
KEYWORDS JP 2000245464-A/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 10562)
AUTHORS Narimatsu,H., Isshiki,S., Togayauchi,A. and Sasaki,K.
TITLE Novel polypeptide
JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;

E38420 10562 bp DNA linear PAT 31-JAN-2002

```
COMMENT
KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464-A/2
PD 12-SEP-2000
PE 25-FEB-1999 JP 1999047571
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAUVAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09.A01867/027.C12N1/21.C12N5/10.C12N9/10.C12P19/00.PC
C12P21/02,
PC C12P21/08.C12Q1/68.G01N33/53/(C12N1/21.C12R1/185).(C12N5/10,
C12R1/91),
PC C12P21/02.C12R1/185).(C12P21/02.C12R1/91).C12N15/00.C12N5/00,
PC {C12N5/00.C12R1/91}
CC
FH Key Location/Qualifiers
FT Promoter (1)..(5000)
FT exon (5001)..(5140)
FT exon (5001)..(5273)
FT exon (5459)..(5567)
FT exon (7427)..(7586)
FT exon (8234)..(10562).
FEATURES
Source
1..10562 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2610 a 2415 c 2574 g 2963 t
ORIGIN
Query Match 100.0%; Score 750; DB 6; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGCCTCTGGCAGGTAGACCTTGAAGGCAAACTGAGTTGAGGTTTGTAGGACGGAAT 60
DB 1 CGCCTCTGGCAGGTAGACCTTGAAGGCAAACTGAGTTGAGGTTTGTAGGACGGAAT 60
QY 61 AATTACTGCTGGGATGACGACATCCCAACCGTTCTGTGAGGACGACGATGATGCC 120
DB 61 AATTACTGCTGGGATGACGACATCCCAACCGTTCTGTGAGGACGACGATGATGCC 120
QY 121 AGTTTGGCAGAGGACGAGGTAGACACGTAAGTCCCTGGCCGTGCTACACCACC 180
DB 121 AGTTTGGCAGAGGACGAGGTAGACACGTAAGTCCCTGGCCGTGCTACACCACC 180
QY 181 ACTGTTTGGAGCTGAGATGTGAACCGAGGCGCTTCGATGCCAAATTCCTCATCTTC 240
DB 181 ACTGTTTGGAGCTGAGATGTGAACCGAGGCGCTTCGATGCCAAATTCCTCATCTTC 240
QY 241 ATCTAGCAGGCTGCTCGGTTAGCAGAGGAGCTCCTGTATCTGCTCTGAGCTTCT 300
DB 241 ATCTAGCAGGCTGCTCGGTTAGCAGAGGAGCTCCTGTATCTGCTCTGAGCTTCT 300
QY 301 TCAGCTGATTTATATGAAACAGAGTAGATATTGTTGGCAATTAGTGAATATATAT 360
DB 301 TCAGCTGATTTATATGAAACAGAGTAGATATTGTTGGCAATTAGTGAATATATAT 360
QY 361 GAGATCATCATAGCAAACTTCACAGTTTGTATCAAGATCCTGCTTCATATCTGCCCA 420
DB 361 GAGATCATCATAGCAAACTTCACAGTTTGTATCAAGATCCTGCTTCATATCTGCCCA 420
QY 421 ACTGATGTGTAAGACAGCTGCAAGAACTTCAGAGCTGCACAAAAAGCAAACTCCAGAC 480
DB 421 ACTGATGTGTAAGACAGCTGCAAGAACTTCAGAGCTGCACAAAAAGCAAACTCCAGAC 480
QY 481 TTATTTCCTGGAATCTGTTTGTGAGACACTGGCCCATGAAATGCTCTCCCAAGAAATAG 540
DB 481 TTATTTCCTGGAATCTGTTTGTGAGACACTGGCCCATGAAATGCTCTCCCAAGAAATAG 540
QY 541 TCGGATTTCTGTCGAATAAATTTGGCAATCTACAGACATGTCTTTTTCAGAGAT 600
DB 541 TCGGATTTCTGTCGAATAAATTTGGGCAATCTACAGACATGTCTTTTTCAGAGAT 600
```

```
QY 601 TTATTTTAAATTAACCTTATTTAAATAATTAACATGCTACAAATTTGCATATAGTGAAGT 660
DB 601 TTATTTTAAATTAACCTTATTTAAATAATTAACATGCTACAAATTTGCATATAGTGAAGT 660
QY 661 GTGCAAAATCTTCGCTACATGGCTCAATGAGTTTTTACATATATTTCCACCATGTAATCA 720
DB 661 GTGCAAAATCTTCGCTACATGGCTCAATGAGTTTTTACATATATTTCCACCATGTAATCA 720
QY 721 CCACCGAGATCTAGAGTAAGATGATTCGT 750
DB 721 CCACCGAGATCTAGAGTAAGATGATTCGT 750
RESULT 2
AF064860 170121 bp DNA linear PRI 05-MAR-2002
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
sequence.
AF064860
ACCESSION AF064860.2 GI:18958624
VERSION HTG: HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170121)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Zimmermann,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Shintani,A., Sasaki,T., Nagamine,K., Kawasaki,K., Asakawa,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Honischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichert,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Resselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
TITLE
JOURNAL
MEDLINE 20289799
PUBMED 10830953
REFERENCE 2 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
TITLE
JOURNAL Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
TITLE
JOURNAL Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
TITLE
JOURNAL Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Feb 27, 2002 this sequence version replaced gi:3171153.
FEATURES Location/Qualifiers
1..170121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
BASE COUNT 45184 a 36756 c 37509 g 50672 t
ORIGIN
```

Query Match 100.0%; Score 750; DB 9; Length 170121; Best Local Similarity 100.0%; Pred. No. 0; Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	CGCCTCTGCGCAAGGTAGACCTTGAAGCAAACTGAGTTGAGTTTGTAGACGGAAT	60						
Db	81815	CGCCTCTGCGCAAGGTAGACCTTGAAGCAAACTGAGTTGAGTTTGTAGACGGAAT	81874						
Qy	61	AATTACTGCTGGGCATGACGACCTTCCCAACCGTTCTGTGAGCGACGCACTTATGCC	120						
Db	81875	AATTACTGCTGGGCATGACGACCTTCCCAACCGTTCTGTGAGCGACGCACTTATGCC	81934						
Qy	121	AGTTTGCACAAGGACAGCGTGTAGACACGTAAGTGCCTGGCCGCTACACCAACC	180						
Db	81935	AGTTTGCACAAGGACAGCGTGTAGACACGTAAGTGCCTGGCCGCTACACCAACC	81994						
Qy	181	ACTGTGTTTTCAGCTGAGATGTGAACAGGCGCTTCTGATTCCTCAATTCCTCATCTTC	240						
Db	81995	ACTGTGTTTTCAGCTGAGATGTGAACAGGCGCTTCTGATTCCTCAATTCCTCATCTTC	82054						
Qy	241	ATCCTACAGCGCTCCCTGGGTTAGCAGAGGGGACTCTGTATCTGTCTGCAGCTTCT	300						
Db	82055	ATCCTACAGCGCTCCCTGGGTTAGCAGAGGGGACTCTGTATCTGTCTGCAGCTTCT	82114						
Qy	301	TCAGCTGATTATTAATGGAAACACAGATAGATATTGATTGGCAATTAGTGAATATTAT	360						
Db	82115	TCAGCTGATTATTAATGGAAACACAGATAGATATTGATTGGCAATTAGTGAATATTAT	82174						
Qy	361	GAGATCATCATGACAACTTCACATTTGATCAAGGATCCTGCCTTCAATATCTGGCCA	420						
Db	82175	GAGATCATCATGACAACTTCACATTTGATCAAGGATCCTGCCTTCAATATCTGGCCA	82234						
Qy	421	ACTGATGTGTAAGAGCAGCTGCAAGAACTTCAGAGCTGACAAAAAGCAAACTCCAGAC	480						
Db	82235	ACTGATGTGTAAGAGCAGCTGCAAGAACTTCAGAGCTGACAAAAAGCAAACTCCAGAC	82294						
Qy	481	TTATTTTCCTGGAATCTGTTTGTGAGACACTGGCCCATGAAATGCTCTCCAGAAATAG	540						
Db	82295	TTATTTTCCTGGAATCTGTTTGTGAGACACTGGCCCATGAAATGCTCTCCAGAAATAG	82354						
Qy	541	TCCGATTTGGTCAATAAATTTGGCAATCTACAGAACATGCTCTTTTTCAGAGAT	600						
Db	82355	TCCGATTTGGTCAATAAATTTGGCAATCTACAGAACATGCTCTTTTTCAGAGAT	82414						
Qy	601	TTATTTTAAATTAACATTTTAAAAATATTAAACATGGTACAAATTTGCATATAGTGAAGT	660						
Db	82415	TTATTTTAAATTAACATTTTAAAAATATTAAACATGGTACAAATTTGCATATAGTGAAGT	82474						
Qy	661	GTGCAATCTTCCTACATGGCTCAATGAGTTTTCATATATATTTCCACCCCATGTAATCA	720						
Db	82475	GTGCAATCTTCCTACATGGCTCAATGAGTTTTCATATATATTTCCACCCCATGTAATCA	82534						
Qy	721	CCACCGAGATCTAGAATAGAATGATTTCGT	750						
Db	82535	CCACCGAGATCTAGAATAGAATGATTTCGT	82564						
RESULT 3									
HS21C080 340000 bp DNA linear PRI 24-MAY-2000									
LOCUS Homo sapiens chromosome 21 segment HS21C080.									
DEFINITION AL163280 AP001735 BA000005									
ACCESSION AL163280.2 GI:7717369									
VERSION									
KEYWORDS									
SOURCE human.									
ORGANISM Homo sapiens									
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
1 (bases 1 to 340000)									
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,									
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,									
Ohki,M., Takeda,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,									
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,									

Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S., Antonarakis,S.E., Minooshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,K. and Yaspo,M.L.	TITLE JOURNAL
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)	COMMENT
The Chromosome 21 Mapping and Sequencing Consortium consists of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagami-hara 228-8555, Japan, * e.mail: sakaki@gscc.riken.go.jp and * URL: http://hgp.gscc.riken.go.jp/	
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/	
* Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8562, Japan, * e.mail: shimizu@mb.med.keio.ac.jp * URL: http://adenine.dmb.med.keio.ac.jp/	
and * GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/	
* Max-Planck Institute for Molecular Genetics, * Innestrasse 73, D-14195 Berlin, Germany, * e.mail: info-chr21@molgen.mpg.de * URL: http://chr21.rz-berlin.mpg.de/.	FEATURES source
Location/Qualifiers 1..340000 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" 1..125945 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone.lib="P160L9, 5' partial" /clone.lib="RPC11,3-5 PAC library" /note="Accession No. AF121897" 73410..243533 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone.lib="P70124" /clone.lib="RPC11,3-5 PAC library" /note="Accession No. AF064860" 224137..>340000 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone.lib="P206A10, 3' partial" /clone.lib="RPC11,3-5 PAC library" /note="Accession No. AF121782" 286528..>340000 /organism="Homo sapiens"	source

Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,
* e.mail: sakaki@gsr.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/ and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/ and
* Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: shimizu@db-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/ and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/ and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info.chr21smolgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/ Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P160L9, 5' partial"
/clone_lib="RPC11.3-5 PAC library"
/notes="Accession No. AF121897"
73410. 243533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11.3-5 PAC library"
/notes="Accession No. AF064860"
224137. >340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11.3-5 PAC library"
/notes="Accession No. AF121897"
286628. >340000
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088..1566
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398..2493
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(2496..2878)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
2879..2931
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2932..3004
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3005..3074
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3473..19157
/gene="SH3BGR"
join(<3473..3565,12649..12678,15331..15460,18739..19157)
/gene="SH3BGR"
join(3473..3565,12649..12678,15331..15426)
/partial
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/db_xref="GI:7717370"
/translation="GSEKAEGGETEAKGESDVGNLPEAQKNEEGGTATEETEE
TAMGAGEAEETETAGEPEGEDEDS"
3473..3565
/gene="SH3BGR"
/number=4
5542..5829
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6887..7067)
/note="L1MD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8482..8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8865..8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
9131..9424
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(9723..9862)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9989..10077)
/note="L2"

```

```

/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10078..10429)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
10980..11000
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(11002..11282)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/gene="SH3BGR"
/number=5
complement(13115..13425)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
/note="(CA)n"
/rpt_family="Simple_repeat"

```

```

Query Match      100.0%; Score 750; DB 9; Length 340000;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCTCTGGCAAGGTAGACCTTGAAGGCAAAACTGAGTTGAGTTGTTAGGACGAAAT 60
Db 155223 CGCCTCTGGCAAGGTAGACCTTGAAGGCAAAACTGAGTTGAGTTGTTAGGACGAAAT 155282

QY 61 AATTACTGTGGCATGCAGCACTTCCCAACCGTTCTGTGAGGAGGAGTGTTATTGCC 120
Db 155283 AATTACTGTGGCATGCAGCACTTCCCAACCGTTCTGTGAGGAGGAGTGTTATTGCC 155342

QY 121 AGTTTGGCAAGGGCACAGGTGTAGAACACGTAAGTGCCTGGGCCCTGTGTACACACC 180
Db 155343 AGTTTGGCAAGGGCACAGGTGTAGAACACGTAAGTGCCTGGGCCCTGTGTACACACC 155402

QY 181 ACTGTGTTTGTAGCTGAGATGTGAACAGGGCCTTCTGTATTCGAATTCCTCATCTTTC 240
Db 155403 ACTGTGTTTGTAGCTGAGATGTGAACAGGGCCTTCTGTATTCGAATTCCTCATCTTTC 155462

QY 241 ATCCTAGCAGGCTGCTGGGTTTAGCAGAAGGGGACTCCTGTATCTGTCTGCAGCTTCT 300
Db 155463 ATCCTAGCAGGCTGCTGGGTTTAGCAGAAGGGGACTCCTGTATCTGTCTGCAGCTTCT 155522

QY 301 TCAGCTGATTTTAAATGGAAACACAGTAGATATTGATTGGCAATTAGTGAATATTAT 360
Db 155523 TCAGCTGATTTTAAATGGAAACACAGTAGATATTGATTGGCAATTAGTGAATATTAT 155582

QY 361 GAGATCATCATAGCAAACTTCACAGTTTGATCAAGGATCCTGCTTCAATATCTGCCA 420
Db 155583 GAGATCATCATAGCAAACTTCACAGTTTGATCAAGGATCCTGCTTCAATATCTGCCA 155642

```

```

QY 421 ACTGATGTGTAAAGCAGCTGCAAGACTTCAGAGCTGACAAAAAAGCAAACTCCAGAC 480
|||||
Db 155643 ACTGATGTGTAAAGCAGCTGCAAGACTTCAGAGCTGACAAAAAAGCAAACTCCAGAC 155702
|||||
QY 481 TTTATTTCTGGAATCTGTTTTGTGAGACACTGGCCCATGAAATGCTCTCCAGAAATAG 540
|||||
Db 155703 TTTATTTCTGGAATCTGTTTTGTGAGACACTGGCCCATGAAATGCTCTCCAGAAATAG 155762
|||||
QY 541 TCGGATTTGGTCAAAATTAATTTGGGCAATTTCTACAGAACATGTGCTTTTTCAGAGAT 600
|||||
Db 155763 TCGGATTTGGTCAAAATTAATTTGGGCAATTTCTACAGAACATGTGCTTTTTCAGAGAT 155822
|||||
QY 601 TTTATTTTAAATTAACCTTTATTTAAAAATATTAAACATGTTACAAATTTGCATATAGTGAAGT 660
|||||
Db 155823 TTTATTTTAAATTAACCTTTATTTAAAAATATTAAACATGTTACAAATTTGCATATAGTGAAGT 155882
|||||
QY 661 GTGCAAACTTCGCTACATGGCTCAATGAGTTTTTACATATATATTTCCACCCCATGTAATCA 720
|||||
Db 155883 GTGCAAACTTCGCTACATGGCTCAATGAGTTTTTACATATATATTTCCACCCCATGTAATCA 155942
|||||
QY 721 CCACCGAGATCTAGATAGATGATTGCT 750
|||||
Db 155943 CCACCGAGATCTAGATAGATGATTGCT 155972
|||||

RESULT 4
AC073231 185982 bp DNA linear HTG 17-JUN-2000
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-729A4, WORKING DRAFT
SEQUENCE, 25 unordered pieces.
ACCESSION
AC073231
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 185982)
AUTHORS
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 185982)
AUTHORS
Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1 1724: contig of 1724 bp in length
1725 1824: gap of unknown length
1825 2971: contig of 1147 bp in length
2972 3071: gap of unknown length
3072 4660: contig of 1589 bp in length
4661 4761: gap of unknown length
4761 5930: contig of 1170 bp in length
5931 6030: gap of unknown length
6031 8127: contig of 2097 bp in length
8128 8228: contig of unknown length
8228 10160: contig of 1933 bp in length
10161 10261: gap of unknown length
10261 13236: contig of 2976 bp in length
13237 16603: contig of unknown length
16604 16703: gap of unknown length

-----
20884: contig of 3981 bp in length
20784: gap of unknown length
24835: contig of 4051 bp in length
24935: gap of unknown length
29727: contig of 4792 bp in length
29827: gap of unknown length
34947: contig of 5120 bp in length
35047: gap of unknown length
41164: contig of 6117 bp in length
41264: gap of unknown length
47050: contig of 5786 bp in length
47150: gap of unknown length
53084: contig of 5934 bp in length
53184: gap of unknown length
60244: contig of 7440 bp in length
60724: gap of unknown length
69057: contig of 8333 bp in length
69157: gap of unknown length
78529: contig of 9372 bp in length
78630: gap of unknown length
85475: contig of 6846 bp in length
85575: gap of unknown length
97151: contig of 11576 bp in length
97251: gap of unknown length
113111: contig of 15860 bp in length
113212 113211: gap of unknown length
129870: contig of 16659 bp in length
129871: gap of unknown length
145368: contig of 15598 bp in length
145569 145668: gap of unknown length
145669 164596: contig of 18928 bp in length
164597 164696: gap of unknown length
164697 185982: contig of 21286 bp in length.

FEATURES
source
1..185982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-729A4"
1..1724
/note="assembly_name:Contig3"
1825..2971
/note="assembly_name:Contig4"
3072..4660
/note="assembly_name:Contig5"
4761..5930
/note="assembly_name:Contig6"
clone_end:17
vector_side:right"
6031..8127
/note="assembly_name:Contig7"
8228..10160
/note="assembly_name:Contig8"
10261..13236
/note="assembly_name:Contig9"
13337..16603
/note="assembly_name:Contig10"
16704..20684
/note="assembly_name:Contig11"
20785..24935
/note="assembly_name:Contig12"
24936..29727
/note="assembly_name:Contig13"
clone_end:sp6
vector_side:right"
29828..34947
/note="assembly_name:Contig14"
35048..41164
/note="assembly_name:Contig15"
41265..47050
/note="assembly_name:Contig16"
47151..53084
/note="assembly_name:Contig17"

```



```
misc_feature 53185..60624
              /note="assembly_name:Contig18"
misc_feature 60725..69057
              /note="assembly_name:Contig19"
misc_feature 69158..78529
              /note="assembly_name:Contig20"
misc_feature 78630..85475
              /note="assembly_name:Contig21"
misc_feature 85576..97151
              /note="assembly_name:Contig22"
misc_feature 97252..113111
              /note="assembly_name:Contig23"
misc_feature 113212..129870
              /note="assembly_name:Contig24"
misc_feature 129971..145568
              /note="assembly_name:Contig25"
misc_feature 145569..164596
              /note="assembly_name:Contig26"
misc_feature 164597..185982
              /note="assembly_name:Contig27"
BASE COUNT 53704 a 39120 c 39561 g 51192 t 2405 others
ORIGIN
Query Match 37.28; Score 279; DB 2; Length 185982;
Best Local Similarity 100.0%; Pred.No.1.2e-150;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 ACTCCAGACTTTATTTCCTGGAATCTGTTTGTGAGACATGCGCCCATGAATGCTCTCC 531
      |||||||
Db 25169 ACTCCAGACTTTATTTCCTGGAATCTGTTTGTGAGACATGCGCCCATGAATGCTCTCC 25228
QY 532 CAGAATAGTCGGATTTGTTGTCGAATTAATTTGGCAATTTCTACAGAACATGTTCTTT 591
      |||||||
Db 25229 CAGAATAGTCGGATTTGTTGTCGAATTAATTTGGCAATTTCTACAGAACATGTTCTTT 25288
QY 592 TTCAGAGATTATTTTAAATAAATTTTAAATAATTAACATGCTACAAATTTGCATA 651
      |||||||
Db 25289 TTCAGAGATTATTTTAAATAAATTTTAAATAATTAACATGCTACAAATTTGCATA 25348
QY 652 TAGTGAAGTGCACAAATTCCTGCATACATGGTCAATGAGTTTACATATATTTCCACC 711
      |||||||
Db 25349 TAGTGAAGTGCACAAATTCCTGCATACATGGTCAATGAGTTTACATATATTTCCACC 25408
QY 712 ATGTAATCACCACCGAGATCTAGATAGAAATGTTCTGT 750
      |||||||
Db 25409 ATGTAATCACCACCGAGATCTAGATAGAAATGTTCTGT 25447

RESULT 5
AC073231/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone RP11-729A4, WORKING DRAFT
SEQUENCE, 25 unordered pieces.
ACCESSION AC073231
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185982)
Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185982)
AUTHORS Waterston,R.H.
DIRECT SUBMISSION
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
```

```
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1724: contig of 1724 bp in length
* 1624: gap of unknown length
* 1725
* 1825
* 2971: contig of 1147 bp in length
* 3071: gap of unknown length
* 3072
* 3072
* 4660: contig of 1589 bp in length
* 4661
* 4760: gap of unknown length
* 4761
* 5930: contig of 1170 bp in length
* 5931
* 6030: gap of unknown length
* 6031
* 8127: contig of 2097 bp in length
* 8128
* 8227: gap of unknown length
* 8228
* 10160: contig of 1933 bp in length
* 10161
* 10260: gap of unknown length
* 10261
* 13236: contig of 2976 bp in length
* 13237
* 13336: gap of unknown length
* 13337
* 16603: contig of 3267 bp in length
* 16604
* 16703: gap of unknown length
* 16704
* 20684: contig of 3981 bp in length
* 20685
* 20784: gap of unknown length
* 20785
* 24835: contig of 4051 bp in length
* 24836
* 24935: gap of unknown length
* 24937
* 29727: contig of 4792 bp in length
* 29728
* 29827: gap of unknown length
* 29828
* 34947: contig of 5120 bp in length
* 34948
* 35047: gap of unknown length
* 35048
* 41164: contig of 6117 bp in length
* 41165
* 41264: gap of unknown length
* 41265
* 47050: contig of 5786 bp in length
* 47051
* 47150: gap of unknown length
* 47151
* 5084: contig of 5934 bp in length
* 5085
* 5184: gap of unknown length
* 5185
* 60624: contig of 7440 bp in length
* 60625
* 60724: gap of unknown length
* 60725
* 69057: contig of 8333 bp in length
* 69058
* 69157: gap of unknown length
* 69158
* 78529: contig of 9372 bp in length
* 78530
* 78629: gap of unknown length
* 78630
* 83475: contig of 8846 bp in length
* 83476
* 85575: gap of unknown length
* 85576
* 97151: contig of 11576 bp in length
* 97152
* 97251: gap of unknown length
* 97252
* 11311: contig of 15860 bp in length
* 11312
* 11311: gap of unknown length
* 11312
* 129870: contig of 16659 bp in length
* 129871
* 129970: gap of unknown length
* 129971
* 145568: contig of 15598 bp in length
* 145569
* 145668: gap of unknown length
* 145669
* 164596: contig of 18928 bp in length
* 164597
* 164696: gap of unknown length
* 164697
* 185982: contig of 21286 bp in length.
FEATURES
          Location/Qualifiers
          1..185982
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="17"
            /clone="RP11-729A4"
          1..1724
            /note="assembly_name:Contig3"
          1825..2971
            /note="assembly_name:Contig4"
          3072..4660
            /note="assembly_name:Contig5"
          4761..5930
            /note="assembly_name:Contig6
clone_end:17
vector_side:right"
6091..8127
            misc_feature
          misc_feature
          misc_feature
          misc_feature
          misc_feature
```

```
misc_feature      /note="assembly_name:Contig7"
6228..10160
/note="assembly_name:Contig8"
10261..13236
/note="assembly_name:Contig9"
13337..16603
/note="assembly_name:Contig10"
16704..20884
/note="assembly_name:Contig11"
20785..24835
/note="assembly_name:Contig12"
24936..29727
/note="assembly_name:Contig13
clone_end:SP6
vector_side:right"
29828..34947
/note="assembly_name:Contig14"
35048..41164
/note="assembly_name:Contig15"
41265..47050
/note="assembly_name:Contig16"
47151..53084
/note="assembly_name:Contig17"
53185..60624
/note="assembly_name:Contig18"
60725..69057
/note="assembly_name:Contig19"
69158..78529
/note="assembly_name:Contig20"
78630..85475
/note="assembly_name:Contig21"
85576..97151
/note="assembly_name:Contig22"
97252..113111
/note="assembly_name:Contig23"
113212..129870
/note="assembly_name:Contig24"
129971..145568
/note="assembly_name:Contig25"
145669..164596
/note="assembly_name:Contig26"
164697..185982
/note="assembly_name:Contig27"
53704 a 39120 c 39561 g 51192 t 2405 others
ORIGIN
Query Match      16.9%; Score 127; DB 2; Length 185982;
Best Local Similarity 100.0%; Pred. No. 5.2e-62;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCTCTGCAAGGTAGACCTTGAAGGCAAACTGAGTTGAGTTGTTAGGACGGAAT 60
|||||
Db 6185 CGCCTCTGCAAGGTAGACCTTGAAGGCAAACTGAGTTGAGTTGTTAGGACGGAAT 6126

QY 61 AATTACTGTGGCATGCAGCACTCCCAACCGTTCTGTAGGACGAGTGTTATTGCC 120
|||||
Db 6125 AATTACTGTGGCATGCAGCACTCCCAACCGTTCTGTAGGACGAGTGTTATTGCC 6066

QY 121 AGTTGG 127
|||||
Db 6065 AGTTGG 6059

RESULT 6
E06819/c
LOCUS      4368 bp      RNA      linear      PAT 29-SEP-1997
DEFINITION DNA encoding mutant of N-methyl-D-aspartate type glutamic acid
            receptor (NMDA type glutamic acid receptor).
ACCESSION  E06819
VERSION     E06819.1 GI:2175001
KEYWORDS   JP 1994062861-A/1.
SOURCE     Mus sp.
ORGANISM   Mus sp.
```

```
REFERENCE
AUTHORS      Mishina,M.
TITLE        NEW PROTEIN AND GENE CODING THE PROTEIN
JOURNAL      MITSUBISHI KASEI CORP
COMMENT      OS Mus sp.(mouse)
            PN JP 1994062861-A/1
            PD 08-MAR-1994
            PF 12-AUG-1992 JP 1992215017
            PI MISHINA MASAMI
            PC C12N15/12,C07K13/00//A61K37/02,C12P21/02;
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            CC *source: tissue=Cerebellum;
            FH Key Location/Qualifiers
            FT CDS 1..4368
            FT /product="mutant of NMDA type glutamic acid
            FT receptor,
            FT epsilon2 subunit"
            FT /note="epsilon2-N589Q".
            FT Location/Qualifiers
            FT 1..4368
            FT /organism="Mus sp."
            FT /db_xref="taxon:10095"
BASE COUNT 1104 a 1187 c 1124 g 953 t
ORIGIN
Query Match      3.2%; Score 24; DB 6; Length 4368;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
|||||
Db 597 GCTCTGCAGCTTCTTCAGCTGATT 574

RESULT 7
I19108/c
LOCUS      4368 bp      DNA      linear      PAT 07-OCT-1996
DEFINITION Sequence 17 from patent US 5502166.
ACCESSION  I19108
VERSION     I19108.1 GI:1599463
KEYWORDS   Unknown.
SOURCE      Unclassified.
            Unknown.
REFERENCE 1 (bases 1 to 4368)
AUTHORS  Mishina,M.
TITLE    NMDH receptor proteins and genes encoding the same
JOURNAL  Patent: US 5502166-A 17 26-MAR-1996;
FEATURES  Location/Qualifiers
            source
            1..4368
            /organism="unknown"
BASE COUNT 1104 a 1187 c 1124 g 953 t
ORIGIN
Query Match      3.2%; Score 24; DB 6; Length 4368;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
|||||
Db 597 GCTCTGCAGCTTCTTCAGCTGATT 574

RESULT 8
E06594/c
LOCUS      4398 bp      RNA      linear      PAT 29-SEP-1997
```

```

DEFINITION cDNA encoding epsilon 2 subunit of type NMDA mouse glutamic acid
receptor.
ACCESSION E06594
VERSION E06594.1 GI:2174781
KEYWORDS JP 1994014783-A/1.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4398)
REFERENCE 1. Mishina, M.
Molecular diversity of the NMDA receptor channel
NATURE 356 (6381), 36-41 (1992)
AUTHORS Araki, K., Meguro, H., Masaki, H., Kumanishi, T., Arakawa, M. and
Mishina, M.
TITLE Molecular diversity of the NMDA receptor channel
JOURNAL NATURE 356 (6381), 36-41 (1992)
MEDLINE 92310364
REFERENCE 2 (bases 1 to 4449)
AUTHORS Kutsuwada, T.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 4449)
AUTHORS Kutsuwada, T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) Tatsuya Kutsuwada, Niigata Univ. Brain
Res. Inst., Dept. of Neuropharmacology; 1-757 Asahimachidori,
Niigata, Niigata 951, Japan (Tel:025-223-6161(ex.5174),
Fax:025-225-6458)
COMMENT Sequence updated (14-Oct-1992) by: Tatsuya Kutsuwada
Submitted (28-FEB-1992) to DBJ by:
Kutsuwada
Department of Neuropharmacology
Brain Research Institute
Niigata University
1-757 Asahimachidori
Niigata 951
Japan
Phone: 025-223-6161 x5147
Fax: 025-225-6458.
Location/Qualifiers
1. 4449
/organism="Mus sp."
/db_xref="taxon:10095"
BASE COUNT 1100 a 1199 c 1137 g 962 t
ORIGIN
Query Match 3.2%; Score 24; DB 6; Length 4398;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
|||||
Db 651 GCTCTGCAGCTTCTTCAGCTGATT 628
|||||

RESULT 9
I19102/c
LOCUS I19102 4446 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 6 from patent US 5502166.
ACCESSION I19102
VERSION I19102.1 GI:1599457
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4446)
AUTHORS Mishina, M.
TITLE NMDA receptor proteins and genes encoding the same
JOURNAL Patent: US 5502166-A 6 26-MAR-1996;
FEATURES
source
Location/Qualifiers
1. 4446
/organism="unknown"
BASE COUNT 1118 a 1208 c 1148 g 972 t
ORIGIN
Query Match 3.2%; Score 24; DB 6; Length 4446;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
|||||
Db 675 GCTCTGCAGCTTCTTCAGCTGATT 652
|||||

```

```

RESULT 10
MUSGRP2/c
LOCUS MUSGRP2/c 4449 bp mRNA linear ROD 02-FEB-1999
DEFINITION Mouse mRNA for glutamate receptor channel subunit epsilon 2.
ACCESSION D10651
VERSION D10651.1 GI:220419
KEYWORDS NMDA receptor; epsilon 2 subunit; glutamate receptor channel.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Kutsuwada, T., Kashiwabuchi, N., Mori, H., Sakimura, K., Kushiya, E.,
Araki, K., Meguro, H., Masaki, H., Kumanishi, T., Arakawa, M. and
Mishina, M.
Molecular diversity of the NMDA receptor channel
NATURE 356 (6381), 36-41 (1992)
92310364
REFERENCE 2 (bases 1 to 4449)
AUTHORS Kutsuwada, T.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 4449)
AUTHORS Kutsuwada, T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1992) Tatsuya Kutsuwada, Niigata Univ. Brain
Res. Inst., Dept. of Neuropharmacology; 1-757 Asahimachidori,
Niigata, Niigata 951, Japan (Tel:025-223-6161(ex.5174),
Fax:025-225-6458)
COMMENT Sequence updated (14-Oct-1992) by: Tatsuya Kutsuwada
Submitted (28-FEB-1992) to DBJ by:
Kutsuwada
Department of Neuropharmacology
Brain Research Institute
Niigata University
1-757 Asahimachidori
Niigata 951
Japan
Phone: 025-223-6161 x5147
Fax: 025-225-6458.
Location/Qualifiers
1. 4449
/organism="Mus musculus"
/db_xref="taxon:10090"
1. 4449
/codon_start=1
/product="glutamate receptor channel subunit epsilon 2"
/protein_id="BAA01498.1"
/db_xref="GI:220420"
/translation="MKPSAECCSPKFWLVLAVLAVSGSKARSKQSPSISGIAVILVGT
SDEVAIKDAHEKDDFHLVSVPRVELVAMNETDPKSIITRICDLMSDRKIOGVVLADD
TDQEAIAQLDFISAQOTLPIGLIHGGSMIMADKDESMFQFGPSIEQOASVMLNI
MEEDWTIFSIYITTFEGYQDFVKNIRSTIENSVGHEVEVLDDMSLDGDSKIQN
QLKQSPIDILLYCTKEEATIFEVANSVGLTYGYTVFVPSVAGDTDVPSEFTG
LISVYDEWDYGLPARVDGIAITTAASDMLSEHSIPEPKSCVNTHEKRIYQNM
LNRYLINVFEGRLNLSFSEDGYQMHPLVILLNKKERKWERVKDKSLQMKYIYVP
RMCPEDEDDHLIVTLEAPFVIVESVDPLSGTCMRNTVPCQKRIISENKTDEEP
GYIKCKGCDILKISKVKETDLYLVTKGKCKKINGTMGIMGVVPMKRAYM
AVGLTINERSEVDFSVFIETGISVWVSRNGTVSPAPLEPSADVWVMFVML
LIVSAVAVFVEFSPVNRCLADGREGPGSFITGKALLWLLGWLFVNNVSPVQPK
GTTSKIMVSWAFPAVIFLASYTANLAAFMIOEYVDVSGLSGLKQFQRPNDPFPFR
FGTVNGSTERNIRNNAEMHAYKQFQNDVDDALLSLTKGLDAFIYDAVLNTMA
GRDEGCKLVITIGSGKVEASTGYIAIOKDSGMKQVDLAILQLFGGEMBEALWLT
GICHNEKNEVMSOLDIDNMAGVFYMLGAAMALSLITFICEHLYFQGRHCFMGVCSG
KPGWFSISRCIYCIHGVALEERQSVNHSPTATMNTNTHSLRLILRTAKMANLSGV
NGSPSALDFIRRESSYDISEHRRSTSDCKSYNNPPCEENLFSDYISEVERTGN
LOLSDSNYQDHYHHHRPHSIGTSSIDGLYDCDNPFTTQPRSIKPKPLDGLPSS
KQSLGLSYGKESFSDRYSHDDLI RSDVSDISTHTVTYGNIEGNAAKRKKQYKDS
LKRPASAKRSRDEFLAYRRRPPRPDKRYFRDKEGLRDFYLDQFQFKEKKNLNV
EHDVLTDIYKRSRDEKRDVSGGGPCTNFKHKGTHGTDKHGVGVGVVPAPWKNLTV
DWEDRSNGNFCRSCPKLHNSSTVAGONSQKQACIRCEACKAGNLYDISDNLQ
LDQPAAPVAVSSNASTTKYFQSPTNRAQKRNKLRKHSYDTFYDLOKEEALAPR
SVSLKDRGRFMDGSPAHMFEMPAGESFANKSVTTAGHHHNNPGSYMLSKSLIPD

```

```

FEATURES
source

```

```

CDS

```

```

RVTONPEIPLGMSACFTAAAPTSGSPRWGRRKQGTSGPLSPISQWCRPFHGAV
PGRFQKIDICIGNOSFCVFNPNKPNRAFGSSNGHYELSSIESDV*
BASE COUNT      1119 a 1208 c 1149 g 973 t
ORIGIN
Query Match      3.2%; Score 24; DB 10; Length 4449;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 GCTCTGCAGCTTCTTCAGCTGATT 310
      |||||||
Db 675 GCTCTGCAGCTTCTTCAGCTGATT 652

RESULT 11
AC124590/c      235126 bp      DNA      linear      HTG 09-JUL-2002
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP23-133P12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC124590
AC124590.2 GI:21717279
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 235126)
McPherson,J.D. and Waterston,R.H.
TITLE
The sequence of Mus musculus clone
JOURNAL
Unpublished
2 (bases 1 to 235126)
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 235126)
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (09-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 9, 2002 this sequence version replaced gi:21426730.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0133P12
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator; 100%
Chemistry: Dye-terminator; 0% of reads
Assembly: Program: Phrap; version 0.990319
Consensus quality: 228361 bases at least Q40
Consensus quality: 230084 bases at least Q30
Consensus quality: 231224 bases at least Q20
Insert size: 210000; agarose-fp
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 8.05 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1462: contig of 1462 bp in length

```

```

* 1463 1562: gap of unknown length
* 1563 3015: contig of 1453 bp in length
* 3016 3115: gap of unknown length
* 3116 4667: contig of 1552 bp in length
* 4668 4767: gap of unknown length
* 4768 7125: contig of 2358 bp in length
* 7126 7225: gap of unknown length
* 7226 13699: contig of 6474 bp in length
* 13700 13799: gap of unknown length
* 13800 25625: contig of 11826 bp in length
* 25626 25725: gap of unknown length
* 25726 36983: contig of 11257 bp in length
* 36983 37082: gap of unknown length
* 37083 58100: contig of 21018 bp in length
* 58101 58200: gap of unknown length
* 58201 86982: contig of 28782 bp in length
* 86983 87082: gap of unknown length
* 87083 128367: contig of 41285 bp in length
* 128368 128467: gap of unknown length
* 128468 177788: contig of 49321 bp in length
* 177789 177888: gap of unknown length
* 177889 233825: contig of 55937 bp in length
* 233826 233925: gap of unknown length
* 233926 235126: contig of 1201 bp in length.

FEATURES
Location/Qualifiers
source
1..235126
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-133P12"
1..1462
/feature="assembly_name:Contig14"
1563..3015
/feature="assembly_name:Contig17"
3116..4667
/feature="assembly_name:Contig18"
4768..7125
/feature="assembly_name:Contig19"
7226..13699
/feature="assembly_name:Contig20"
13800..25625
/feature="assembly_name:Contig21"
25726..36982
/feature="assembly_name:Contig22"
37083..58100
/feature="assembly_name:Contig23"
58201..86982
/feature="assembly_name:Contig24"
87083..128367
/feature="assembly_name:Contig25"
128468..177788
/feature="assembly_name:Contig26"
177889..233825
/feature="assembly_name:Contig27"
233926..235126
/feature="assembly_name:Contig6"

BASE COUNT 66400 a 50894 c 50677 g 1206 others
ORIGIN

Query Match      3.2%; Score 24; DB 2; Length 235126;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 GCTCTGCAGCTTCTTCAGCTGATT 310
      |||||||
Db 77526 GCTCTGCAGCTTCTTCAGCTGATT 77503

RESULT 12
AC090786
LOCUS
DEFINITION
Homo sapiens chromosome 8 clone RP11-381G11 map 8, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.

```

```

AC090786
AC090786.4 GI:22474932
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168407)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-381G11
Unpublished
2 (bases 1 to 168407)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Boguslavsky,L., Bouckghalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 168407)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckghalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 24, 2002 this sequence version replaced gi:18642764.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12774
Center clone name: 381_G_11
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 61343: contig of 61343 bp in length
 * 61344 61443: gap of 100 bp
 * 61444 141596: contig of 80153 bp in length
 * 141597 141696: gap of 100 bp
 * 141697 168407: contig of 26711 bp in length.

FEATURES
 source
 1..168407
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="8"
 /chromosome="8"
 /clone="RP11-381G11"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 50359 a 37373 c 34544 g 45879 t 252 others
 ORIGIN

Query Match 3.1%; Score 23; DB 2; Length 168407;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 AGTTGATCAAGGATCCTGCCIT 407
 |||||||||||||||||||||||||||
 Db 50365 AGTTGATCAAGGATCCTGCCIT 50387

RESULT 13
 AC007628 140356 bp DNA linear PRI 30-JUN-1999
 LOCUS
 Genomic sequence of Homo sapiens clone N0576M10 from chromosome 18,
 complete sequence.
 AC007628
 AC007628.3 GI:5296051
 HTG.
 SOURCE
 Homo sapiens.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 140356)
 McCombie,W.R., Haberman,K., Gnoj,L., de la Bastide,M., Dedhia,N.N.,
 Matero,A., Ning Huang,E., O'Shaughnessy,A., Preston,R.,
 Rodriguez,M., Schutz,K., Shah,R., Shekher,M., Spiegei,L., Swady,I.
 and Vill,D.
 Genomic Sequence of clone N0576M10 from chromosome 18
 Unpublished
 2 (bases 1 to 140356)
 McCombie,W.R.
 Direct Submission
 Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing
 Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
 Harbor, NY 11724, USA
 3 (bases 1 to 140356)
 McCombie,W.R.
 Direct Submission
 Submitted (30-JUN-1999) Lita Annenberg Hazen Genome Sequencing
 Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
 Harbor, NY 11724, USA
 On Jun 30, 1999 this sequence version replaced gi:5103792.
 N0576M10 overlaps with clone 367D17 (AC003971). The overlap starts
 at base 134980. From bases 120370-120405 and from 120540-120690,
 the assembly of N0576M10 was single stranded/single chemistry.

FEATURES
 source
 1..140356
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /clone="N0576M10"
 BASE COUNT 41499 a 30302 c 29689 g 38866 t

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

FEATURES
 source

BASE COUNT

ORIGIN

Query Match 2.9%; Score 22; DB 9; Length 140356;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 610 ATTAACTATTATAAATATT 631
 |||||
 Db 132868 ATTAACTATTATAAATATT 132889

RESULT 14
 AC009758/c
 LOCUS
 DEFINITION Homo sapiens chromosome 11, clone RP11-101E4, complete sequence.
 AC009758
 AC009758.8 GI:18201851
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 141485)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,D., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Rile,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission
 Submitted (31-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 141485)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
 Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,B., Schupback,R., Seaman,S.,
 Severy,P., Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Theodore,J., Travers,M., Subramanian,A., Talamas,J., Tesfaye,S.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (26-MAY-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 141485)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

```

repeat_region /rpt_family="L1PA13" complement(13735..13897)
repeat_region /rpt_family="MLT1" complement(13917..14126)
repeat_region /rpt_family="MER7A" 14127..14251
repeat_region /rpt_family="L1PA13" complement(14252..14343)
repeat_region /rpt_family="MER7A" 14344..14478
repeat_region complement(16636..16806)
repeat_region /rpt_family="MER3" 17135..17192
repeat_region /rpt_family="AT-rich" 17193..17282
repeat_region /rpt_family="MIR" 17530..17697
repeat_region /rpt_family="MIR" 17775..18110
repeat_region complement(19001..20135)
repeat_region /rpt_family="L2" 20136..20172
repeat_region /rpt_family="TG1n" complement(20173..20263)
repeat_region /rpt_family="L2" 20366..20776
repeat_region /rpt_family="L2" 21814..21893
repeat_region /rpt_family="L2" complement(23678..23873)
repeat_region /rpt_family="L1MEC" complement(24101..24125)
repeat_region /rpt_family="L2" complement(24126..24472)
repeat_region /rpt_family="THE1C" complement(24473..24830)
repeat_region /rpt_family="L2" complement(25722..26067)
repeat_region /rpt_family="MLT1B" complement(26070..26882)
repeat_region /rpt_family="MLT1B-int" complement(26883..27318)
repeat_region /rpt_family="L1MA4" complement(27319..27836)
repeat_region /rpt_family="MLT1B-int" complement(27838..28212)
repeat_region /rpt_family="MLT1B" 28638..28826
repeat_region /rpt_family="L3" complement(30711..30944)
repeat_region /rpt_family="MER3" 31658..31745
repeat_region /rpt_family="MIR" complement(31803..32089)
repeat_region /rpt_family="AluY" 32327..32582
repeat_region /rpt_family="L2" complement(33440..34705)
repeat_region /rpt_family="L2" 34805..35174
repeat_region /rpt_family="L2" 35543..35626
repeat_region /rpt_family="L2" complement(37510..37926)
repeat_region /rpt_family="MSTC" complement(37934..38963)
repeat_region /rpt_family="MSTC-int"

Query Match 2.9% Score 22; DB 9; Length 141485;
Best Local Similarity 100.0%; Pred.No. 0.89;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 238 TTCATCCTAGCAGGCTGCCTGC 259

```

```

Db 33560 TTCATCCTAGCAGGCTGCCTGC 33539

RESULT 15
LOCUS AC079759/c
DEFINITION Homo sapiens chromosome UNK clone RPI1-115M2, *** SEQUENCING IN
PROGRESS *** 51 unordered pieces.
ACCESSION AC079759 147318 bp DNA linear
VERSION AC079759.1 GI:10047960
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147318)
Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 147318)
Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (10-SEP-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 809: contig of 809 bp in length
* 810 909: gap of unknown length
* 910 2541: contig of 1632 bp in length
* 2542 2642: gap of unknown length
* 2642 3783: contig of 1142 bp in length
* 3783 3883: gap of unknown length
* 3883 5585: contig of 1702 bp in length
* 5585 5686: gap of unknown length
* 5686 7052: contig of 1367 bp in length
* 7052 7152: gap of unknown length
* 7152 8501: contig of 1449 bp in length
* 8501 8701: gap of unknown length
* 8701 10222: contig of 1521 bp in length
* 10222 10322: gap of unknown length
* 10322 11546: contig of 1224 bp in length
* 11546 11646: gap of unknown length
* 11646 13518: contig of 1872 bp in length
* 13518 13619: gap of unknown length
* 13619 15318: contig of 1700 bp in length
* 15318 15419: gap of unknown length
* 15419 16954: contig of 1536 bp in length
* 16954 17054: gap of unknown length
* 17054 18815: contig of 1761 bp in length
* 18815 18915: gap of unknown length
* 18915 20379: contig of 1464 bp in length
* 20379 20479: gap of unknown length
* 20479 22421: contig of 1942 bp in length
* 22421 22521: gap of unknown length
* 22521 24698: contig of 2177 bp in length
* 24698 24798: gap of unknown length
* 24798 26997: contig of 2199 bp in length
* 26997 27097: gap of unknown length
* 27097 28411: contig of 1314 bp in length
* 28411 27098

```

```

* 28412 28511: gap of unknown length
* 28512 28712: contig of 1201 bp in length
* 28713 28612: gap of unknown length
* 28813 31906: contig of 2094 bp in length
* 31907 32006: gap of unknown length
* 32007 33712: contig of 1706 bp in length
* 33713 33812: gap of unknown length
* 33813 36109: contig of 2297 bp in length
* 36110 36209: gap of unknown length
* 36210 38530: contig of 2321 bp in length
* 38531 38530: gap of unknown length
* 38631 41289: contig of 2659 bp in length
* 41290 41389: gap of unknown length
* 41390 44082: contig of 2693 bp in length
* 44083 44182: gap of unknown length
* 44183 47171: contig of 2989 bp in length
* 47172 47271: gap of unknown length
* 47272 48708: contig of 1437 bp in length
* 48709 48808: gap of unknown length
* 48809 51230: contig of 2422 bp in length
* 51231 51330: gap of unknown length
* 51331 54121: contig of 2791 bp in length
* 54122 54221: gap of unknown length
* 54222 56131: contig of 1910 bp in length
* 56132 56231: gap of unknown length
* 56232 58592: contig of 2361 bp in length
* 58593 58992: gap of unknown length
* 58993 60612: contig of 1920 bp in length
* 60613 60712: gap of unknown length
* 60713 63636: contig of 2924 bp in length
* 63637 63736: gap of unknown length
* 63737 66096: contig of 2360 bp in length
* 66097 66196: gap of unknown length
* 66197 68503: contig of 2307 bp in length
* 68504 68603: gap of unknown length
* 68604 71387: contig of 2784 bp in length
* 71388 71487: gap of unknown length
* 71488 73973: contig of 2486 bp in length
* 73974 74073: gap of unknown length
* 74074 77332: contig of 3259 bp in length
* 77333 77432: gap of unknown length
* 77433 80459: contig of 3027 bp in length
* 80460 80559: gap of unknown length
* 80560 83571: contig of 3012 bp in length
* 83572 83671: gap of unknown length
* 83672 86856: contig of 3185 bp in length
* 86857 86956: gap of unknown length
* 86957 90657: contig of 3701 bp in length
* 90658 90757: gap of unknown length
* 90758 94182: contig of 3425 bp in length
* 94183 94282: gap of unknown length
* 94283 96909: contig of 2627 bp in length
* 96910 97010: gap of unknown length
* 97010 101791: contig of 4782 bp in length
* 101792 101891: gap of unknown length
* 101892 104889: contig of 2998 bp in length
* 104890 104989: gap of unknown length
* 104990 108599: contig of 3710 bp in length
* 108600 108799: gap of unknown length
* 108800 114115: contig of 5316 bp in length
* 114116 114215: gap of unknown length
* 114216 120675: contig of 6460 bp in length
* 120676 120775: gap of unknown length
* 120776 128360: contig of 7485 bp in length
* 128361 128360: gap of unknown length
* 128361 135922: contig of 7562 bp in length
* 135923 136022: gap of unknown length
* 136023 147318: contig of 11296 bp in length.

```

Location/Qualifiers
1. .147318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"

FEATURES
source

```

misc_feature /clone="Rp11-115M2"
misc_feature 1. .809
misc_feature /note="assembly_name:Contig28"
misc_feature 910. .2541
misc_feature /note="assembly_name:Contig39"
misc_feature 2642. .3783
misc_feature /note="assembly_name:Contig41"
misc_feature 3884. .5585
misc_feature /note="assembly_name:Contig43"
misc_feature 5686. .7052
misc_feature /note="assembly_name:Contig44"
misc_feature 7133. .8601
misc_feature /note="assembly_name:Contig46"
misc_feature 8702. .10222
misc_feature /note="assembly_name:Contig47"
misc_feature 10323. .11546
misc_feature /note="assembly_name:Contig48"
misc_feature 11647. .13518
misc_feature /note="assembly_name:Contig49"
misc_feature 13619. .15318
misc_feature /note="assembly_name:Contig50"
misc_feature 15419. .16954
misc_feature /note="assembly_name:Contig52"
misc_feature 17055. .18815
misc_feature /note="assembly_name:Contig53"
misc_feature 18916. .20379
misc_feature /note="assembly_name:Contig54"
misc_feature 20480. .22421
misc_feature /note="assembly_name:Contig55"
misc_feature 22522. .24698
misc_feature /note="assembly_name:Contig57"
misc_feature 24799. .26997
misc_feature /note="assembly_name:Contig58"
misc_feature 27098. .28411
misc_feature /note="assembly_name:Contig59"
misc_feature 28512. .29712
misc_feature /note="assembly_name:Contig60"
misc_feature 29813. .31906
misc_feature /note="assembly_name:Contig61"
misc_feature 32007. .33712
misc_feature /note="assembly_name:Contig62"
misc_feature 33813. .36109
misc_feature /note="assembly_name:Contig63"
misc_feature 36210. .38530
misc_feature /note="assembly_name:Contig64"
misc_feature 38631. .41289
misc_feature /note="assembly_name:Contig65"
misc_feature 41390. .44082
misc_feature /note="assembly_name:Contig66"
misc_feature 44183. .47171
misc_feature /note="assembly_name:Contig67"
misc_feature 47272. .48708
misc_feature /note="assembly_name:Contig68"
misc_feature 48809. .51230
misc_feature /note="assembly_name:Contig69"
misc_feature 51331. .54121
misc_feature /note="assembly_name:Contig70"
misc_feature 54222. .56131
misc_feature /note="assembly_name:Contig71"
misc_feature 56232. .58592

```

Query Match 2.9%; Score 22; DB 2; Length 147318;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 238 TTCATCCTAGCAGGCTGCTGC 259
|||||
DB 124514 TTCATCCTAGCAGGCTGCTGC 124493

Search completed: April 12, 2003, 07:30:51
Job time : 2202.77 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 11, 2003, 22:38:14 ; Search time 123.127 Seconds
(without alignments)
13717.504 Million cell updates/sec

Title: US-09-914-152-3_COPY_1_750

Perfect score: 750

Sequence: 1 cgctctggcaagtagacc.....ctagaataagaatgattcgt 750

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750	100.0	10562	21 AAA93876	Human beta3Gal-T5
2	35	4.7	807	20 AAZ16242	Human gene express
3	24	3.2	4368	15 AAQ55979	NMDA receptor chan
4	24	3.2	4446	15 AAQ56916	Glutamic acid rece
5	21	2.8	831	22 AAH65977	C glutamic codin
6	21	2.8	954	22 AAF71244	Corynebacterium gl
7	21	2.8	954	22 AAF67882	Corynebacterium gl
8	21	2.8	6167	24 ABK28370	DNA transcription
9	21	2.8	349980	22 AAH68527	C glutamic codin

10	20	2.7	6535	23 ABL11862	Drosophila melanog
11	20	2.7	19597	22 AAK78483	Human immune/haema
12	19	2.5	441	23 AAS55482	Streptococcus pneu
13	19	2.5	441	23 AAS55840	Streptococcus pneu
14	19	2.5	476	24 ABN65139	Human cancer relat
15	19	2.5	813	19 AAR98739	DNA encoding a S.
16	19	2.5	846	24 ABQ89886	Human prostate exp
17	19	2.5	1530	18 AAT68657	Thermococcus beta-
18	19	2.5	1530	19 AAV36910	Thermococcus 9N2-3
19	19	2.5	6679	19 AAV65233	DNA encoding RNA p
20	19	2.5	6914	15 AAQ55757	Enterobacter cloac
21	19	2.5	7189	24 ABN80027	Human chemically m
22	19	2.5	8752	19 AAY31189	E. coli J96 pathog
23	19	2.5	8991	23 ABL29234	Drosophila melanog
24	19	2.5	9118	24 ABQ66990	Human angiogenesis
25	19	2.5	9416	22 ABA89519	Escherichia coli p
26	19	2.5	9543	24 ABN80049	Human chemically m
27	19	2.5	11831	19 AAV52198	Streptococcus pneu
28	19	2.5	12437	23 ABL17704	Drosophila melanog
29	19	2.5	117574	24 AAL45288	Human KCNB1 gene.
30	19	2.5	2944528	24 ABA03041	Listeria monocytog
31	18	2.4	127	16 AAT26779	Human gene signatu
32	18	2.4	228	20 AAH86056	Human single nucle
33	18	2.4	522	22 AAL18459	Human breast cance
34	18	2.4	615	20 AAX95959	Nucleic acid sequ
35	18	2.4	622	24 ABK93475	Human breast speci
36	18	2.4	710	22 AAH98518	Human EST-derived
37	18	2.4	843	24 ABQ32626	Oligonucleotide fo
38	18	2.4	843	24 ABQ32627	Oligonucleotide fo
39	18	2.4	865	19 AAV48123	Nucleotide sequenc
40	18	2.4	865	21 AAC66206	Human galectin 11
41	18	2.4	865	21 AAZ49417	Human galectin 11
42	18	2.4	1330	21 AAC66208	Human galectin 11b
43	18	2.4	1337	21 AAC66207	Human galectin 11a
44	18	2.4	1375	22 AAD06493	Arabidopsis thalia
45	18	2.4	1643	21 AAZ50940	Human extracellular

ALIGNMENTS

RESULT 1
AAA93876
ID AAA93876 standard; DNA; 10562 BP.

AC AAA93876;

XX 15-JAN-2001 (first entry)

XX Human beta3Gal-T5 encoding DNA.

DE Human beta3Gal-T5 encoding DNA.
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
digestive system; beta3Gal-T5; ds.

OS Homo sapiens.

XX WO2000050508-A1.

XX 31-AUG-2000.

XX 24-FEB-2000; 2000WO-JP01070.

XX 25-FEB-1999; 99JP-0047571.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;

XX WPI; 2000-549409/50.

PT Beta-1,3 galactose transferase and DNA encoding it, useful for
synthesis of type I sialyl Lewis x, a carbohydrate for treatment of
digestive system cancer

XX PS Claim 31; Page 103-111; 123pp; Japanese.

XX CC This invention relates to a polypeptide (I) with beta-1,3 galactose

CC transferase activity, or variants of (I) comprising amino acid additions,

CC deletions and/or substitutions. Included in the invention is DNA encoding

CC all or part of (I); expression vectors containing the DNA, host cells

CC transformed by the vectors; a method for the preparation of the

CC polypeptide by culture of the transformants or by expression in the milk

CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3

CC galactose transferase protein transfers galactose by beta-1,3 bonding to

CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as

CC GlcNAc) to give Galbeta1-4Glc. The protein and

CC DNA encoding it are useful for the treatment and diagnosis of cancer of

CC the digestive system. The present sequence represents a Beta3gal-T5

XX CC encoding DNA sequence.

XX SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 100.0%; Score 750; DB 21; Length 10562;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGCTCTGGCAAGGTAGACCTTGAAGGCAAACTGAGTTGAGTTGTAGACGGAAT	60
Db	1	CGCTCTGGCAAGGTAGACCTTGAAGGCAAACTGAGTTGAGTTGTAGACGGAAT	60
Qy	61	AATTAATCTGCTGGGATGACACCTTCCCAACCGTTCTGTAGGCGAGGAGTGTATTGCC	120
Db	61	AATTAATCTGCTGGGATGACACCTTCCCAACCGTTCTGTAGGCGAGGAGTGTATTGCC	120
Qy	121	AGTTGGCACAGGGCACAGGTGTAGAACACGTAAAGTGCCTGGCGGTGTACACACC	180
Db	121	AGTTGGCACAGGGCACAGGTGTAGAACACGTAAAGTGCCTGGCGGTGTACACACC	180
Qy	181	ACTGTGTTGAGTGTGAGATGTGAACGAGGCGCTTCGTATCCAAATTCCTCATTCCTTC	240
Db	181	ACTGTGTTGAGTGTGAGATGTGAACGAGGCGCTTCGTATCCAAATTCCTCATTCCTTC	240
Qy	241	ATCTAGCAGGTCGCTCGGTTAGCAGAGGGAGTCTGCTGTATCTGCTCGACCTCT	300
Db	241	ATCTAGCAGGTCGCTCGGTTAGCAGAGGGAGTCTGCTGTATCTGCTCGACCTCT	300
Qy	301	TCAGCTGATTTAATGAAACAGAGTAGATATTGATTGGCAATTAGTGAATATTAT	360
Db	301	TCAGCTGATTTAATGAAACAGAGTAGATATTGATTGGCAATTAGTGAATATTAT	360
Qy	361	GAGATCATCATCACTCACTTCAGTTTATCAAGATCCCTGCCTCAATATCTGGCCA	420
Db	361	GAGATCATCATCACTCACTTCAGTTTATCAAGATCCCTGCCTCAATATCTGGCCA	420
Qy	421	ACTGATGTGTTAAAGCAGCTGCAAGAACTTCAGAGCTGACAAAAAGCAAACTCCAGAC	480
Db	421	ACTGATGTGTTAAAGCAGCTGCAAGAACTTCAGAGCTGACAAAAAGCAAACTCCAGAC	480
Qy	481	TTTATTCCTGGAATCTGTTTGTGACACTGGCCCATGAATGCTCCCGAAGAAATAG	540
Db	481	TTTATTCCTGGAATCTGTTTGTGACACTGGCCCATGAATGCTCCCGAAGAAATAG	540
Qy	541	TCGATTTGGTCAATAAATTGGGCAATTCACAGAACATGTGTCTTTTCAGAGAT	600
Db	541	TCGATTTGGTCAATAAATTGGGCAATTCACAGAACATGTGTCTTTTCAGAGAT	600
Qy	601	TTATTTTAATTAACCTATTAAATATTAACATGGTACAAATTCGATATAGTGAAGT	660
Db	601	TTATTTTAATTAACCTATTAAATATTAACATGGTACAAATTCGATATAGTGAAGT	660
Qy	661	GTGCAATCTTCGCTACATGCTCAATGAGTTTTTACATATATTTCCACCCATGTAATCA	720
Db	661	GTGCAATCTTCGCTACATGCTCAATGAGTTTTTACATATATTTCCACCCATGTAATCA	720
Qy	721	CCACGAGATCTAGAAATGATGATTCGT	750
Db	721	CCACGAGATCTAGAAATGATGATTCGT	750

Db 721 CCACGAGATCTAGAAATGATGATTCGT 750

RESULT 2

AAZ16242

ID AAZ16242 standard; cDNA; 807 BP.

XX AAZ16242;

XX 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:3712.

XX Human: gene; gene expression product; diagnosis; therapy; probe;

XX detection; mapping; tissue typing; profiling; forensic; cancer;

XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX W09938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX Jones W, Kassam A, Kennedy GC, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1; Page 1770; 2479pp; English.

XX The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ1779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

XX SQ Sequence 807 BP; 188 A; 187 C; 190 G; 203 T; 39 other;

Query Match 4.7%; Score 35; DB 20; Length 807;

Best Local Similarity 100.0%; Pred. No. 2.5e-07;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 TTGTTAGGACGAAATAATTACTGCTGGCATGCA 79
 DB 469 TTGTTAGGACGAAATAATTACTGCTGGCATGCA 503

RESULT 3
 AAQ55979/c
 ID AAQ55979 standard; DNA; 4368 BP.
 XX
 AC AAQ55979;
 XX
 DT 20-SEP-1994 (first entry)
 XX
 DE NMDA receptor channel subunit epsilon-2-N589Q.
 XX
 KW NMDA; N-methyl-D-aspartate; receptor; channel; subunit;
 KW brain; cerebellum; neurotransmission; synapse; memory;
 KW learning; neurocyte; necrosis; cerebral ischemia;
 KW status epilepticus; CNS; central nervous system; ss.
 XX
 OS Mus musculus.
 XX
 XX Key Location/Qualifiers
 FT misc_difference 1785..1777
 FT /*tag= a
 FT /note= "Asn589 codon replaced by Gln codon"
 XX
 PN W09404698-A.
 XX
 PD 03-MAR-1994.
 XX
 PF 12-AUG-1993; 93WO-JP01143.
 XX
 PR 12-AUG-1992; 92JP-0215017.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 XX
 PI Mishina M;
 XX
 DR WPI; 1994-083212/10.
 DR P-PSDB; AAR49042.
 XX
 PT NMDA receptor channel epsilon and zeta sub-unit proteins -
 PT obtained by Xenopus oocyte expression of modified receptor mRNA
 XX
 PS Claim 1; Page 11-20; 44pp; Japanese.
 XX
 CC NMDA receptor cDNA from a mouse cerebellum-derived library
 CC is subjected to site-specific mutagenesis to give cDNA coding
 CC for the desired modified proteins. mRNA derived from this
 CC is then expressed in Xenopus laevis oocytes.
 CC The modified NMDA receptor channel subunit proteins (AAQ55979-81)
 CC are useful as tools for investigation of the mechanism of
 CC neurotransmission across synapses, the development of synapse
 CC plasticity (basic to memory and learning), the development of
 CC neurocyte necrosis as a result of disorders such as cerebral
 CC ischemia and status epilepticus. This is necessary for understanding
 CC the mechanisms of neurotransmission in the CNS and the organisation
 CC and pathology of the brain, and in the development of drugs and
 CC therapy for brain diseases and genetically determined disorders.
 XX
 SQ Sequence 4368 BP; 1104 A; 1187 C; 1124 G; 953 T; 0 other;

Query Match 3.2%; Score 24; DB 15; Length 4368;
 Best Local Similarity 100.0%; Pred. NO. 0.086;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
 DB 597 GCTCTGCAGCTTCTTCAGCTGATT 574

RESULT 5
 AAH65977
 ID AAH65977 standard; DNA; 831 BP.
 XX
 AC AAH65977;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 1012.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.

RESULT 4
 AAQ56916/c
 ID AAQ56916 standard; DNA; 4446 BP.
 XX
 AC AAQ56916;
 XX
 DT 08-AUG-1994. (first entry)
 XX
 DE Glutamic acid receptor.
 XX
 KW Glutamic acid receptor; synaptic signal translation; diagnosis;
 KW brain disease; nerve signal; gene therapy; NMDA; cerebellum; ICR;
 KW mouse; nerve cell necrosis; ds.
 XX
 OS Mus musculus.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..4446
 FT /*tag= a
 FT /product= glutamic_acid_receptor
 XX
 PN JP06014783-A.
 XX
 PD 25-JAN-1994.
 XX
 PF 30-JUN-1992; 92JP-0173155.
 XX
 PR 30-JUN-1992; 92JP-0173155.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 XX
 DR WPI; 1994-061478/08.
 DR N-PSDB; AAQ56916.
 XX
 PT New glutamic acid receptor and gene - for use in analysis of
 PT synaptic signal translation, and diagnosis of brain disease
 XX
 PS Claim 1; Page 32-35; 35pp; Japanese.
 XX
 CC The sequence codes for a glutamic acid receptor. The receptor is
 CC useful for the analysis of nerve signal translation; within the
 CC synapse, expression of synapse plasticity, nerve cell necrosis,
 CC brain structure and brain disease. It can also be used in gene
 CC therapy.
 XX
 SQ Sequence 4446 BP; 1118 A; 1208 C; 1148 G; 972 T; 0 other;

Query Match 3.2%; Score 24; DB 15; Length 4446;
 Best Local Similarity 100.0%; Pred. NO. 0.086;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
 DB 675 GCTCTGCAGCTTCTTCAGCTGATT 652

RESULT 5
 AAH65977
 ID AAH65977 standard; DNA; 831 BP.
 XX
 AC AAH65977;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 1012.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.

```
XX PD 20-JUN-2001.
XX PF
XX PR 18-DEC-2000; 2000EP-0127588.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PR (KYOW ) KYOWA HAKKO KOGYO KK.
XX PA
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX PR
XX DR WPI; 2001-376931/40.
XX DR P-PSDB; AAG90758.
XX PR
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS
XX PS Claim 8; SEQ ID NO: 1012; 246pp + Sequence Listing; English.
XX CC
XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a nucleic acid described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.
XX SQ Sequence 831 BP; 162 A; 162 C; 287 G; 220 T; 0 other;

Query Match 2.8%; Score 21; DB 22; Length 831;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 ATATTGATTGGCAATTAGTG 351
Db 404 ATATTGATTGGCAATTAGTG 424

RESULT 6
AAF71244
ID AAF71244 standard; DNA; 954 BP.
XX AC
XX AC AAF71244;
XX DT
XX DT 30-APR-2001 (first entry)
XX DE
XX DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:213.
XX KW
XX KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
XX KW fine chemical production; organic acid; proteinoogenic amino acid;
XX KW nonproteinoogenic amino acid; purine base; pyrimidine base; nucleoside;
XX KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
XX KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
XX KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
XX KW Brevibacterium; environmental condition; ds.
XX OS
XX OS Corynebacterium glutamicum.
XX PN W0200100842-A2.
XX XX
XX PD 04-JAN-2001.
XX XX
```

```
PF 23-JUN-2000; 2000WO-IB00911.
XX PR
XX PR 25-JUN-1999; 99US-0141031.
XX PR 08-JUL-1999; 99DE-1031636.
XX PR 09-JUL-1999; 99DE-1031235.
XX PR 09-JUL-1999; 99DE-1031236.
XX PR 09-JUL-1999; 99DE-1031237.
XX PR 09-JUL-1999; 99DE-1031238.
XX PR 09-JUL-1999; 99DE-1031239.
XX PR 09-JUL-1999; 99DE-1032226.
XX PR 14-JUL-1999; 99DE-1032920.
XX PR 14-JUL-1999; 99DE-1032922.
XX PR 14-JUL-1999; 99DE-1032924.
XX PR 14-JUL-1999; 99DE-1032928.
XX PR 14-JUL-1999; 99DE-1032930.
XX PR 14-JUL-1999; 99DE-1032933.
XX PR 14-JUL-1999; 99DE-1032935.
XX PR 14-JUL-1999; 99DE-1032973.
XX PR 14-JUL-1999; 99DE-1033002.
XX PR 14-JUL-1999; 99DE-1033003.
XX PR 14-JUL-1999; 99DE-1033005.
XX PR 14-JUL-1999; 99DE-1033006.
XX PR 31-AUG-1999; 99DE-1041378.
XX PR 31-AUG-1999; 99DE-1041379.
XX PR 31-AUG-1999; 99DE-1041390.
XX PR 31-AUG-1999; 99DE-1041391.
XX PR 03-SEP-1999; 99DE-1042088.
XX PA (BADI ) BASF AG.
XX PI
XX PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX DR
XX DR WPI; 2001-061974/07.
XX DR P-PSDB; AAB79129.
XX PT
XX PT New isolated Corynebacterium glutamicum nucleic acid for production or
XX PT modulation of production of fine chemicals such as amino acids,
XX PT nucleosides, lipids, fatty acids, carbohydrates, vitamins
XX PT or enzymes .
XX PS
XX PS Claim 3; Page 428-429; 712pp; English.
XX CC
XX CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
XX CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
XX CC C. glutamicum HA genes (I) can be used in vectors for expression in host
XX CC cells and production of fine chemicals, such as, an organic acid,
XX CC proteinoogenic or nonproteinoogenic amino acid (preferred), purine or
XX CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
XX CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
XX CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
XX CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
XX CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
XX CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
XX CC be modulated. The presence of (I) or HA proteins encoded by then are
XX CC used for diagnosing the presence or activity of Corynebacterium
XX CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
XX CC used as markers for genetically engineered Corynebacterium or
XX CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
XX CC homeostasis in C. glutamicum or help the microorganism to adapt to
XX CC different environmental conditions.
XX SQ Sequence 954 BP; 189 A; 187 C; 316 G; 262 T; 0 other;

Query Match 2.8%; Score 21; DB 22; Length 954;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 ATATTGATTGGCAATTAGTG 351
Db 504 ATATTGATTGGCAATTAGTG 524

RESULT 7
```

AAAF67882
ID AAFF67882 standard; DNA; 954 BP.
XX
XX
AC AAF67882;
XX
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:279.
XX
XX
KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
XX
PN W0200100805-A2.
XX
XX
PD 04-JAN-2001.
XX
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
XX
25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 08-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
XX
(BADI) BASF AG.
XX
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX
WPI; 2001-071486/08.
DR
DR
P-PSDB; AAB76649.
XX
XX
Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
XX
Claim 3; Page 559-560; 1119pp; English.
PS
XX

CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Warner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 6167 BP; 1658 A; 282 C; 1483 G; 2741 T; 3 other;
 Query Match 2.8%; Score 21; DB 24; Length 6167;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 613 AAACCTATTATAAAATATTAA 633
 Db 3207 AAACCTATTATAAAATATTAA 3187

RESULT 9
 AAH68527
 ID AAH68527 standard; DNA; 349980 BP.
 AC AAH68527;
 DT 26-SEP-2001 (first entry)
 XX C glutamicum coding sequence fragment SEQ ID NO: 7062.
 DE Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX Corynebacterium glutamicum.
 OS
 XX EP1108790-A2.
 PN
 XX 20-JUN-2001.
 PD
 XX 18-DEC-2000; 2000EP-0127688.
 PF
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280986.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI; 2001-376931/40.
 DR
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX Disclosure; SEQ ID NO: 7062; 246pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium, and identifying a homologue of a gene derived
 CC from Corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 349980 BP; 79725 A; 90426 C; 98918 G; 80911 T; 0 other;
 Query Match 2.8%; Score 21; DB 22; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 ATATTGATTGGCAATTAGTG 351
 Db 63212 ATATTGATTGGCAATTAGTG 63232

RESULT 10
 ABL11862
 ID ABL11862 standard; cDNA; 6535 BP.
 XX ABL11862;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30068.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB67759.
 DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 PT
 XX Claim 1; SEQ ID NO 30068; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
 CC sequences (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 6535 BP; 1923 A; 1344 C; 1352 G; 1916 T; 0 other;
 Query Match 2.7%; Score 20; DB 23; Length 6535;
 Query Match

Best Local Similarity 100.0%; Pred. No. 8.8;		Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	604 TTTTAAATTAACATATTTA 623		
Db	5639 TTTTAAATTAACATATTTA 5658		
RESULT 11			
ID	AAK78483		
AC	AAK78483 standard; DNA; 19597 BP.		
AC	AAK78483;		
DT	07-NOV-2001 (first entry)		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33295.		
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;		
KW	cytostatic; gene therapy; vaccine; metastasis; ds.		
OS	Homo sapiens.		
PN	WO200157182-A2.		
PD	09-AUG-2001.		
XX	17-JAN-2001; 2001WO-US01354.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205535.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225577.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226581.		
PR	22-AUG-2000; 2000US-0226866.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	02-OCT-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	13-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	20-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241221.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		
PR	08-NOV-2000; 2000US-0246475.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246478.		
PR	08-NOV-2000; 2000US-0246523.		
PR	08-NOV-2000; 2000US-0246524.		
PR	08-NOV-2000; 2000US-0246525.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246527.		
PR	08-NOV-2000; 2000US-0246528.		
PR	08-NOV-2000; 2000US-0246532.		
PR	08-NOV-2000; 2000US-0246533.		
PR	08-NOV-2000; 2000US-0246609.		
PR	08-NOV-2000; 2000US-0246610.		
PR	08-NOV-2000; 2000US-0246611.		
PR	08-NOV-2000; 2000US-0246613.		
PR	17-NOV-2000; 2000US-0249207.		
PR	17-NOV-2000; 2000US-0249208.		
PR	17-NOV-2000; 2000US-0249209.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249211.		
PR	17-NOV-2000; 2000US-0249212.		
PR	17-NOV-2000; 2000US-0249213.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249215.		
PR	17-NOV-2000; 2000US-0249216.		
PR	17-NOV-2000; 2000US-0249217.		
PR	17-NOV-2000; 2000US-0249218.		
PR	17-NOV-2000; 2000US-0249244.		

```

PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX PI WPI: 2001-483426/52.
XX DR
XX PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating cancers and
XX PT metastasis.
XX PS Disclosure; SEQ ID NO 33295; 3071pp + Sequence Listing; English.
XX CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
XX CC activity, and can be used in gene therapy and vaccine production. (I)
XX CC proteins and polynucleotides may be used in the prevention, diagnosis and
XX CC treatment of diseases associated with inappropriate (I) expression. For
XX CC example, they may be used to treat disorders associated with decreased
XX CC expression by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of (I) by expressing inactive proteins or to
XX CC supplement the patients' own production of (I). Additionally, (I)
XX CC polynucleotides may be used to produce the secreted (I).
XX CC The nucleic acids into a host cell and culturing the cell to express the
XX CC protein. (I) proteins and polynucleotides may be used to prevent,
XX CC diagnose and treat immune/hematopoietic-related diseases, especially
XX CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX CC to AAK87694 represent human immune/hematopoietic antigen genomic
XX CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX CC represent sequences used in the exemplification of the present invention.
XX SQ Sequence 19597 BP; 6593 A; 2910 C; 3794 G; 6300 T; 0 other;

Query Match 2.7%; Score 20; DB 22; Length 19597;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TGGTCAAAATAATTGGGCA 569
DB 2909 TGGTCAAAATAATTGGGCA 2928
|||||

RESULT 12
AAS55482/C
ID AAS55482 standard; DNA; 441 BP.
XX AC AAS55482;
XX DT 13-FEB-2002 (first entry)
XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #53.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.

XX OS Streptococcus pneumoniae.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX PI P-PSDB; AAU37623.
XX DR WPI: 2001-611495/70.
XX DR
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Claim 27; Seq ID No 9119; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes, themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 441 BP; 138 A; 79 C; 104 G; 120 T; 0 other;

Query Match 2.5%; Score 19; DB 23; Length 441;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGCTGATTA 312
DB 405 AGCTTCTTCAGCTGATTA 387
|||||

RESULT 13
AAS55840/C
ID AAS55840 standard; DNA; 441 BP.
XX AC AAS55840;
XX DT 13-FEB-2002 (first entry)
XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #411.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.

```



```
XX OS Streptococcus pneumoniae.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-20727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253623P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR P-PSDB; AAU37981.
XX PS New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX CC Claim 27; Seq ID No 9477; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence encodes an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 441 BP; 138 A; 79 C; 104 G; 120 T; 0 other;

Query Match 2.5%; Score 19; DB 23; Length 441;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGCTGATTGA 312
DB 405 AGCTTCTTCAGCTGATTGA 387
|||||
RESULT 14
ABN65139
ID ABN65139 standard; CDNA; 476 BP.
XX AC ABN65139;
XX AC ABN65139;
XX DT 28-JUN-2002 (first entry)
XX DE Human cancer related polynucleotide SEQ ID NO 5106.
XX KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
gene therapy; cancer; tumour; gene; ss.
```

```
XX OS Homo sapiens.
XX PN WO200214500-A2.
XX PD 21-FEB-2002.
XX PF 16-AUG-2001; 2001WO-US25840.
XX PR 16-AUG-2000; 2000US-226326P.
XX PA (CHIR) CHIRON CORP.
XX PA (HYSE-) HYSEQ INC.
XX PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
XX PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX DR WPI; 2002-241905/29.
XX PT New nucleic acid for producing a polypeptide, detecting differentially
XX PT expressed genes correlated with a cancerous state of a mammalian cell,
XX PT and inhibiting tumor growth -
XX PS Claim 1; SEQ ID NO 5106; 883pp + Sequence Listing; English.
XX CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
XX CC with cytostatic activity. The polynucleotide is used to produce a
XX CC polypeptide, to detect differentially expressed genes correlated with a
XX CC cancerous state of a mammalian cell and to inhibit tumour growth. The
XX CC polynucleotide is used as a probe in mapping and tissue profiling. The
XX CC encoded polypeptide and antibodies to the polypeptide can also be used
XX CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
XX CC gene therapy.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 476 BP; 135 A; 68 C; 102 G; 171 T; 0 other;

Query Match 2.5%; Score 19; DB 24; Length 476;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GGTCAAAATAAATTGGGCA 569
DB 243 GGTCAAAATAAATTGGGCA 261
|||||
RESULT 15
AAT98739/c
ID AAT98739 standard; DNA; 813 BP.
XX AC AAT98739;
XX AC AAT98739;
XX DT 09-NOV-1998 (first entry)
XX DE DNA encoding a S. pneumoniae 50S ribosomal protein L15.
XX KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
immunological response; inoculation; antibody production; inhibitor;
XX KW T cell immune response; antimicrobial compound; bacterial adhesion;
XX KW extracellular matrix protein; protein-mediated cell invasion; wound;
pathogenesis; ss.
XX OS Streptococcus pneumoniae.
XX PH Key Location/Qualifiers
XX FT CDS 230..670
XX FT /*tag= a
XX PN WO9743303-A1.
XX PD 20-NOV-1997.
```

```
XX 14-MAY-1997; 97WO-US07950.
XX
XX 14-MAY-1996; 96US-0017670.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
XX Stodola RK;
XX
XX WPI; 1998-008793/01.
XX P-PSDB; AAW38697.
XX
XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
XX diagnosing anti-microbial agents for treatment of bacterial
XX infections
XX
XX Claim 4; Page 231; 483pp; English.
XX
XX This sequence encodes a Streptococcus pneumoniae protein that (based on
XX homology with a Bacillus stearothermophilus protein) is a 50S ribosomal
XX protein 15, and represents a DNA sequence of the invention.
XX The DNA sequences were isolated from Streptococcus pneumoniae strain
XX 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
XX invention can be used to identify compounds which interact with and
XX inhibit or activate the activity of the proteins. Antagonists can be
XX used to treat diseases caused by S. pneumoniae proteins, through genetic
XX immunisation. They can also be used to induce an immunological response
XX in a mammal by inoculation with the S. pneumoniae proteins or delivery
XX of the encoding nucleic acids in a vector adequate to produce antibody
XX and/or T cell immune responses to protect the animal from disease. The
XX proteins can also be used to identify antimicrobial compounds which are
XX capable of inhibiting their bioactivity. In particular the proteins of
XX the invention can be used to prevent adhesion of bacteria to mammalian
XX extracellular matrix proteins on in-dwelling devices or in wounds, to
XX block protein-mediated mammalian cell invasion, and to block the normal
XX progression of pathogenesis in infections initiated other than by the
XX implantation of in-dwelling devices or other surgical techniques.
XX
XX Sequence 813 BP; 253 A; 141 C; 181 G; 238 T; 0 other;
```

```
Query Match 2.5%; Score 19; DB 19; Length 813;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 294 AGCTTCTTCAGCTGATTTA 312
   |||||
Db 534 AGCTTCTTCAGCTGATTTA 616
```

Search completed: April 12, 2003, 03:53:22
Job time : 240.127 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 02:50:35 ; Search time 823.332 seconds
(without alignments)
14752.993 Million cell updates/sec

Title: US-09-914-152-3_COPY_1_750

Perfect score: 750

Sequence: 1 cgctctggcaggtagacc.....ctagaatagaagtattcgt 750

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hcc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hcc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rtd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	3.2	462	10 AW248757	AW248757 2820823.3
c 2	24	3.2	2462	11 AK002963	AK002963 Mus muscu
c 3	21	2.8	176	9 AA787280	AA787280 nla10a1.r
c 4	21	2.8	245	17 CNS03JLD	AU247018 Tetraodon
c 5	21	2.8	283	14 T40667	T40667 ya08c05.sl
c 6	21	2.8	305	17 A2266648	A2266648 RPCI-23-1

7	21	2.8	601	17	AQ581246	AQ581246 RPCI-11-4
c 8	21	2.8	643	17	AZ284360	AZ284360 RPCI-23-1
c 9	20	2.7	109	14	BQ596648	BQ596648 PEESToab2
c 10	20	2.7	363	14	N55969	N55969 J4829F Huma
c 11	20	2.7	420	9	A1076906	A1076906 OY80910.x
c 12	20	2.7	429	9	A1959457	A1959457 fd10d06.y
c 13	20	2.7	430	12	BF326949	BF326949 OV3-BN004
c 14	20	2.7	434	13	BI979420	BI979420 ft87b12.y
c 15	20	2.7	458	13	BI979113	BI979113 ft87b12.x
c 16	20	2.7	545	17	BH607402	BH607402 BOCMO80TR
c 17	20	2.7	548	17	AQ441014	AQ441014 HS_5109.A
c 18	20	2.7	549	17	BH774294	BH774294 uzmb003f0
c 19	20	2.7	551	9	A1288124	A1288124 qv88b02.x
c 20	20	2.7	559	17	AZ412895	AZ412895 1N0186E10
c 21	20	2.7	586	17	BH152926	BH152926 Gm_UMB001
c 22	20	2.7	652	9	AA253932	AA253932 mw07C05.f
c 23	20	2.7	652	17	BH518234	BH518234 BOGEF94TF
c 24	20	2.7	673	10	AV733363	AV733363 AV733363
c 25	20	2.7	678	13	BM269429	BM269429 MEST109-C
c 26	20	2.7	728	17	BH728622	BH728622 BOMLO36TR
c 27	20	2.7	734	17	BH464599	BH464599 BOHOV96TR
c 28	20	2.7	775	10	AW116491	AW116491 fl15603.x
c 29	20	2.7	821	9	AL528186	AL528186 AL528186
c 30	20	2.7	919	12	BG333647	BG333647 602460536
c 31	20	2.7	1100	14	BQ053357	BQ053357 AGENCOURT
c 32	19	2.5	143	17	BH316645	BH316645 CH230-131
c 33	19	2.5	159	10	BB455819	BB455819 BB455819
c 34	19	2.5	195	13	BM431624	BM431624 lduo25A10
c 35	19	2.5	206	10	BB073135	BB073135 BB073135
c 36	19	2.5	237	9	AV239893	AV239893 AV239893
c 37	19	2.5	250	9	AA991537	AA991537 os59f05.s
c 38	19	2.5	272	10	AV362531	AV362531 AV362531
c 39	19	2.5	287	9	A1696665	A1696665 wc55d10.x
c 40	19	2.5	293	10	AW484923	AW484923 63129 MAR
c 41	19	2.5	305	9	A1370544	A1370544 ta39e01.x
c 42	19	2.5	305	9	AV045607	AV045607 AV045607
c 43	19	2.5	327	9	AI823489	AI823489 wh54f11.x
c 44	19	2.5	343	14	H12427	H12427 yj1le12.sl
c 45	19	2.5	357	12	BF510404	BF510404 UT-H-B14-

ALIGNMENTS

RESULT 1
AW248757
LOCUS
DEFINITION
2820823.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2820823 3',
mRNA sequence.
ACCESSION
AW248757
VERSION
AW248757.1 GI:6591750
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 462)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT
Other ESTs: 2820823.5prime
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF cDNA Library Preparation: Ling
Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
Project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bhrp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:

http://www.genome.washington.edu Polyadenylation: Based upon the presence of a xhoi site followed by a run of 14 or more T residues at the beginning of the sequence, this cDNA insert was polyadenylated.

Plate: L1CM5 row: D column: 8
High quality sequence stop: 383.

FEATURES

Source
Location/Qualifiers
1..462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2820823"
/clone_lib="NH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
126 a 110 C 82 g 144 t

BASE COUNT

Query Match 3.2%; Score 24; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. NO. 0.092;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 604 TTTTAACTTAACTTATTAAAAA 627

Db 347 TTTTAACTTAACTTATTAAAAA 370

RESULT 2

AK002963/c

LOCUS

DEFINITION Mus musculus adult male brain cDNA, RIKEN full-length enriched library, clone:071000LJ21:glutamate receptor, ionotropic, NMDA2B (epsilon 2), full insert sequence.

ACCESSION AK002963.1 GI:12833327

VERSION AK002963.1

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male brain cDNA to mRNA,

clone lib:RIKEN full-length enriched mouse cDNA library

CLONE:071000LJ21.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning

JOURNAL High Enzymol. 303, 19-44 (1999)

99279253

10349636

2

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3

Shibata,K., Itoh,M., Aizawa,K., Nagacka,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

TITLE

sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

4

Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kado,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wyshak-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kotsuki,S.,
and Hayashizaki,Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

5 (bases 1 to 2462)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGCGCCGCACTCGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using triazole thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to R₀ = 5.0. Second strand cDNA
was prepared with the primer adaptor of sequence[5'
GAGAGAGAGAGGATCCAGAGCTCATTTAATTAATTAACCCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:

XhoI. Host: SOUR.

Location/Qualifiers

1..2462
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:071000LJ21"
/db_xref="MGI:1896807"
/db_xref="taxon:10090"

FEATURES source

```

/clone="0710001J21"
/sex="male"
/tissue_type="brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1. .2462
/gene="Grin2b"
765. .1778
/gene="Grin2b"
/note="data source:MGD, source key:MGI:95821,
evidence:ISS
putative"
/codon_start=1
/protein_id="BAB22483.1"
/db_xref="GI:12833328"
/db_xref="MGD:MGI:95821"
/translators="MPSAECSPKFWLVLAVSGSKARSQKSPSITGIAVLVGT
SDVAIKDAHEKDDFHLSVPRVELAMNEDPKSIITRICDMSDRKIQGVFADD
TDGEAIAQLDFTISAQTLPTILGHHGSSMINADKDESMFFQFGPSP1EQOASVMLNI
MEYDWIFISIVTYPPGYQDFVFNKIRSTIENSFGWELEELVLLDMSLDGDSKION
OLAKLOSPIILLYCTKEATYIFEVANSVGLGYGTWIVPSLVAGDDETVPSEPTG
LISVSDWDYGLPARVDGIALITTAASDMLSERSFIPERASSYNTHREKRIYOSNM
LNR"
polyA_signal 2438. .2443
/gene="Grin2b"
/note="putative"
2462
/gene="Grin2b"
/note="putative"
polyA_site 534 a 666 c 573 g 589 t
BASE COUNT 534 a 666 c 573 g 589 t
ORIGIN
Query Match 3.2%; Score 24; DB 11; Length 2462;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGAGCTTCTTCAGCTGATT 310
|||||
Db 1439 GCTCTGAGCTTCTTCAGCTGATT 1416

RESULT 3
AA787280/c 176 bp mRNA linear EST 31-JUL-1998
LOCUS nial0a1.r1 Aspergillus nidulans 24hr asexual developmental and
DEFINITION vegetative cDNA lambda zap library Emericella nidulans cDNA clone
nial0a1, mRNA sequence.
ACCESSION AA787280
VERSION AA787280.1 GI:2847511
KEYWORDS EST
SOURCE Emericella nidulans.
ORGANISM Emericella nidulans.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
REFERENCE 1 (bases 1 to 176)
AUTHORS Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
TITLE An Aspergillus nidulans EST Database
JOURNAL Unpublished (1998)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
High quality sequence stop: 140.
FEATURES
source
1. .176
/organism="Emericella nidulans"

/strain="EGSC A26"
/db_xref="taxon:162425"
/clone="nial0a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/tissue_type="vegetative mycelia, asexual structures"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 46 a 36 c 61 g 33 t
ORIGIN
Query Match 2.8%; Score 21; DB 9; Length 176;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CTCGAGCTTCTTCAGCTGATT 310
|||||
Db 63 CTCGAGCTTCTTCAGCTGATT 43

RESULT 4
CNS03JLD 245 bp DNA linear GSS 17-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone
DEFINITION 031L05 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL247018
VERSION AL247018.1 GI:7968030
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 245)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 245)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Barnot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 245)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1. .245
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="031L05"
/note="Genoscope"
/note="Genoscope sequence ID : C0BG031CF03SP1-end ;
pUC-Ori"
BASE COUNT 81 a 34 c 39 g 66 t 25 others
ORIGIN
Query Match 2.8%; Score 21; DB 17; Length 245;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 599 ATTATTTTAAATAAACTTA 619
Db 17 ATTATTTTAAATAAACTTA 37

RESULT 5
T40667/c
LOCUS
DEFINITION
  T40667
  IMAGE:60872 3', mRNA sequence.
ACCESSION
  T40667.1 GI:648270
VERSION
  EST.
KEYWORDS
  human.
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 283)
  Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
  Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
  ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
  ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
  Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
  Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
  Generation and analysis of 280,000 human expressed sequence tags
  Genome Res. 6 (9), 807-828 (1996)
  97044478
Other_ESTRs: ya08c05.r2
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -21m13.

FEATURES
  source
    1..283
    /organism="Homo sapiens"
    /db_xref="GDB:488681"
    /db_xref="taxon:9606"
    /clone="IMAGE:60872"
    /clone_lib="Stratagene placenta (#37225)"
    /sex="male"
    /lab_host="SOLR cells (kanamycin resistant)"
    /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
    ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
    Oligo dt. Caucasian. Average insert size: 1.2 kb; Uni-ZAP
    XR Vector; -5' adaptor sequence: 5' GAATTCGCGACAG 3' -3',
    adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 84 a 35 c 41 g 123 t
ORIGIN
  Query Match 2.8%; Score 21; DB 14; Length 283;
  Best Local Similarity 100.0%; Pred. No. 4.1;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 609 AATTAACCTATTAAAAATA 629
Db 51 AATTAACCTATTAAAAATA 31

RESULT 6
A2266648/c
LOCUS
DEFINITION
  A2266648
  RPCI-23-143C8.TJ RPCI-23 Mus musculus genomic clone RPCI-23-143C8,
  DNA sequence.
ACCESSION
  A2266648
  A2266648.1 GI:9479911
VERSION
  GSS.
KEYWORDS
  house mouse.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 305)
  Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akintret
  ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
  and Fraser,C.M.
  Mouse BAC End Sequences from Library RPCI-23
  Unpublished (1999)
  Other_GSSs: RPCI-23-143C8.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 143 row: C column: 8
Seq primer: SP6
Class: BAC ends.

FEATURES
  Location/Qualifiers
    1..305
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPCI-23-143C8"
    /clone_lib="RPCI-23"
    /sex="Female"
    /lab_host="DH10B"
    /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
    ECORI; Site_2: ECORI; Female C57BL/6J mouse kidney and/or
    brain genomic DNA was isolated and partially digested
    with a combination of ECORI and ECORI Methyase. Size
    selected DNA was cloned into the pBAC3.6 vector at the
    ECORI sites. The ligation products were transformed into
    DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 114 a 60 c 39 g 92 t
ORIGIN
  Query Match 2.8%; Score 21; DB 17; Length 305;
  Best Local Similarity 100.0%; Pred. No. 4.1;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 611 TTAACCTATTAAAAATATT 631
Db 186 TTAACCTATTAAAAATATT 166

RESULT 7
AQ581246
LOCUS
DEFINITION
  AQ581246
  RPCI-11-450B12.TV RPCI-11 Homo sapiens genomic clone RPCI-11-450B12
  , DNA sequence.
ACCESSION
  AQ581246
  AQ581246.1 GI:5008356
VERSION
  GSS.
KEYWORDS
  human.
SOURCE
  human.
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 601)
  Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter
  ,J.C.
  Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
  Map Building
  Unpublished (1997)
  Other_GSSs: RPCI-11-450B12.TJ

```

Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1. .601
 /organism="Homo sapiens"
 /db_xref="GDB:7672451"
 /db_xref="taxon:9606"
 /clone="RPCI-11-450B12"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 173 a 105 c 79 g 244 t
 ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 601;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 597 AGATTATTTTAACTAACT 617
 |||||
 Db 127 AGATTATTTTAACTAACT 147

RESULT 8
 A2284360/c 643 bp DNA linear GSS 27-JUL-2000
 LOCUS
 DEFINITION
 RPCI-23-125H16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-125H16
 DNA sequence.

ACCESSION
 A2284360
 VERSION
 A2284360.1 GI:9526146
 KEYWORDS
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus

REFERENCE
 AUTHORS
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 643)
 Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
 ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.

TITLE
 JOURNAL
 COMMENT
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-125H16.TV

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 125 row: H column: 16
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1. .643
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-125H16"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methyase. Size
 selected DNA was cloned into the pBACE3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 168 a 152 c 119 g 204 t
 ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 643;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 612 TAACTTTTAAAAAATTA 632
 |||||
 Db 463 TAACTTTTAAAAAATTA 443

RESULT 9
 BQ596648/c 109 bp mRNA linear EST 24-JUN-2002
 LOCUS
 DEFINITION
 PESTOab20f09.yi Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 falciparum CONA 5', mRNA sequence.
 ACCESSION
 BQ596648
 VERSION
 BQ596648.1 GI:21543374
 KEYWORDS
 EST.
 SOURCE
 malaria parasite P. falciparum.

ORGANISM
 Plasmodium falciparum
 REFERENCE
 1 (bases 1 to 109)
 AUTHORS
 Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
 Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
 Tsagarishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
 Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
 Waterston,R., Wilson,R. and Sibley,D.

TITLE
 JOURNAL
 COMMENT
 WashU Plasmodium EST Project
 Unpublished (2001)
 Contact: L. David Sibley
 WashU Plasmodium EST Project
 Washington University School
 of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library was constructed by Debopam Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: L. David Sibley
 (sibley@porcim.wustl.edu), Washington University
 Seq primer: -40UP from Gibco.

FEATURES

Location/Qualifiers
 1. .109
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /note="Vector: pBlueScript SK plus; Site_1: EcoRI; Site_2:
 XhoI; Library was constructed by Debopam Chakrabarti.
 Total RNA samples were isolated from mixed stage
 saponin(0.1%) -lysed P. falciparum 3D7 infected
 erythrocytes by the acidic guanidinium-phenol chloroform
 method. The poly A+ RNA was isolated by the polyA+ tract
 mRNA isolation system (Promega, WI) using streptavidin

S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

WashU zebrafish EST Project 1998

Unpublished (1998)

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Matthew Clark, cDNA Library Arrayed by:

Matthew Clark, DNA Sequencing by: Washington University Genome

Sequencing Center Clone Distribution: Genome Systems, St. Louis,

Misouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: huf@resgen.com) and

Ressourcenzentrum Primatardatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T3 ET from Amersham.

Location/Qualifiers

1..429

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="IMAGE:3730475"

/clone_lib="Zebrafish WashU MPIMG EST"

/sex="mixed"

/tissue_type="26 somite embryos, adult livers, shield

stage embryos"

/lab_host="XLI-blue MRF"

/note="Vector: pSPOR1; Site_1: NotI; Site_2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dT)15 primer

(5'-GACTAGTCTGATCGAGCGCCGCTTTTGTGTTT3'); double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPOR1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss), adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

BASE COUNT 128 a 103 c 105 g 93 t

ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 429;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CTGCAGCTCTCTCAGCTCAT 309

Db 115 CTGCAGCTCTCTCAGCTCAT 96

RESULT 13

BF326949

LOCUS

BF326949 430 bp mRNA linear EST 22-NOV-2000

DEFINITION QV3-BN0046-270700-282-905 BN0046 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF326949

VERSION BF326949.1

KEYWORDS GI:11297697

EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 430)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l-QV3&t2-QV3-BN0046-

270700-282-905&t3=2000-07-27&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 14

High quality sequence stop: 430.

Location/Qualifiers

1..430

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="BN0046"

/dev_stage="Adult"

/note="Organ: breast; normal; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORFES PCR (O.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 155 a 68 c 54 g 153 t

ORIGIN

Query Match 2.7%; Score 20; DB 12; Length 430;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 ATTAACCTTATTATAAATA 629

Db 93 ATTAACCTTATTATAAATA 112

RESULT 14

BI979420/c

LOCUS

BI979420 434 bp mRNA linear EST 24-OCT-2001

DEFINITION ft87b12.yl Gong zebrafish ovary Danio rerio cDNA clone 5160166 5,

similar to TR:Q9WUL0 Q9WUL0 DNA TOPOISOMERASE I.; mRNA sequence.

ACCESSION BI979420

VERSION BI979420.1

KEYWORDS GI:16363591

SOURCE zebrafish.

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 434)

AUTHORS

Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy

, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood

, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.

and Wilson, R.

WashU zebrafish EST Project 1998

Unpublished (1998)

TITLE

WashU zebrafish EST Project 1998

JOURNAL

COMMENT

Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu
The library was constructed by Dr. Z. Gong. DNA Sequencing by:
Washington University Genome Sequencing Center St. Louis. Please
contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T3 ET from Amersham
High quality sequence stop: 394.

FEATURES

source

1. .434
Location/Qualifiers
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5160166"
/clone_lib="Gong zebrafish ovary"
/sex="female"
/dev_stage="4-5 month"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary (pooled); Vector: pBluescript SK-;
Site1: XhoI; Site2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
mass-excised to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."

143 a 110 c 109 g 72 t
BASE COUNT
ORIGIN

Query Match 2.7%; Score 20; DB 13; Length 434;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CTGCAGCTTCTTCAGCTGAT 309

|||||
Db 183 CTGCAGCTTCTTCAGCTGAT 164

RESULT 15

BI979113

LOCUS BI979113 458 bp mRNA linear EST 24-OCT-2001
DEFINITION ft87b12.x1 Gong zebrafish ovary Danio rerio cDNA clone 5160166 3'
similar to TR:Q9WUL0 Q9WUL0 DNA TOPOISOMERASE I.; mRNA sequence.

BI979113

BI979113.1 GI:16360149

EST.

zebrafish.

ORGANISM

Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.

1 (bases 1 to 458)

REFERENCE
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.

TITLE WashU zebrafish EST Project 1998

JOURNAL Unpublished (1998)

COMMENT Other_ESTs: ft87b12.y1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@wustl.wustl.edu

The library was constructed by Dr. Z. Gong. DNA Sequencing by:
Washington University Genome Sequencing Center St. Louis. Please
contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T7 from Gibco.

FEATURES

source

1. .458
Location/Qualifiers
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5160166"
/clone_lib="Gong zebrafish ovary"
/sex="female"
/dev_stage="4-5 month"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary (pooled); Vector: pBluescript SK-;
Site1: XhoI; Site2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
mass-excised to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."

88 a 115 c 109 g 146 t

BASE COUNT

ORIGIN

Query Match 2.7%; Score 20; DB 13; Length 458;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CTGCAGCTTCTTCAGCTGAT 309

|||||
Db 317 CTGCAGCTTCTTCAGCTGAT 336

Search completed: April 12, 2003, 08:58:58

Job time : 831.332 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 03:59:25 ; Search time 20.8169 Seconds
(without alignments)
11049.076 Million cell updates/sec

Title: US-09-914-152-3_COPY_1_750

Perfect score: 750

Sequence: 1 cgcctctggcaggtagacc.....ctagaataagaatgtattcgt 750

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2.6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2.6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2.6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2.6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2.6/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	24	3.2	4368	1	US-08-026-138E-17
C 2	24	3.2	4446	1	US-08-026-138E-6
C 3	19	2.5	56	1	US-08-139-862-9
C 4	19	2.5	60	1	US-08-139-862-10
C 5	19	2.5	813	4	US-08-858-207A-205
C 6	19	2.5	1330	4	US-09-134-078-4
C 7	19	2.5	6914	1	US-08-920-812-22
C 8	19	2.5	6914	1	US-08-920-827-22
C 9	19	2.5	6914	1	US-08-921-177-22
C 10	19	2.5	6914	1	US-08-362-577C-22
C 11	19	2.5	6914	2	US-08-920-828-22
C 12	19	2.5	8752	4	US-08-976-259-3
C 13	19	2.5	11831	4	US-08-961-527-65
C 14	18	2.4	1843	4	US-09-131-648-3
C 15	18	2.4	1970	1	US-08-028-463-14
C 16	18	2.4	1370	1	US-08-461-836-14
C 17	18	2.4	3645	2	US-08-663-112-1
C 18	18	2.4	4463	2	US-08-760-489-3
C 19	18	2.4	4463	2	US-08-760-489-3
C 20	18	2.4	4463	4	US-09-185-373-1
C 21	18	2.4	4463	4	US-09-185-373-3
C 22	17	2.3	549	3	US-09-109-204-25
C 23	17	2.3	647	3	US-09-109-204-26
C 24	17	2.3	900	4	US-09-221-017B-357
C 25	17	2.3	2224	3	US-09-109-204-6
C 26	17	2.3	2818	1	US-08-366-276-1
C 27	17	2.3	2990	4	US-09-009-816-19

28	17	2.3	5852	1	US-07-867-106-2	Sequence 2, Appli
29	17	2.3	43676	3	US-09-356-952-12	Sequence 12, Appl
C 30	17	2.3	56516	2	US-08-996-306-1	Sequence 1, Appli
C 31	17	2.3	56516	4	US-09-338-907-1	Sequence 1, Appli
C 32	17	2.3	56516	4	US-09-218-207-1	Sequence 1, Appli
C 33	17	2.3	56520	4	US-09-338-907-179	Sequence 179, App
C 34	17	2.3	56520	4	US-09-218-207-179	Sequence 179, App
C 35	17	2.3	152331	3	US-09-128-155-16	Sequence 16, Appl
C 36	17	2.3	176373	3	US-09-128-155-17	Sequence 17, Appl
C 37	17	2.3	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 38	17	2.3	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 39	17	2.3	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 40	16	2.1	28	1	US-08-052-157-6	Sequence 6, Appli
C 41	16	2.1	431	3	US-09-026-343-26	Sequence 26, Appl
C 42	16	2.1	431	4	US-09-362-871-26	Sequence 26, Appl
C 43	16	2.1	455	3	US-09-026-343-20	Sequence 20, Appl
C 44	16	2.1	455	4	US-09-362-871-20	Sequence 20, Appl
C 45	16	2.1	628	1	US-08-686-878A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-026-138E-17/c
: Sequence 17, Application US/08026138E
: Patent No. 5502166
: GENERAL INFORMATION:
: APPLICANT: Masayoshi MISHINA
: TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nishiohata Residence 1-107
: STREET: 5214, Nishiohata-machi
: CITY: Niigata-shi
: STATE: Niigata-ken
: COUNTRY: JAPAN
: ZIP: 951

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS v.5

SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/026,138E
FILING DATE: 26-FEB-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 39563/1992

FILING DATE: 26-FEB-1992
APPLICATION NUMBER: JP 173155/1992

FILING DATE: 30-JUN-1992
APPLICATION NUMBER: JP 215017/1992

FILING DATE: 12-AUG-1992
APPLICATION NUMBER: JP 303878/1992

FILING DATE: 13-NOV-1992
ATTORNEY/AGENT INFORMATION:

NAME: Hamburg, C.Bruce
REGISTRATION NUMBER: 22,389

REFERENCE/DOCKET NUMBER: F-4551
TELEPHONE: (212) 986-2340

TELECOMMUNICATION INFORMATION:
TELEFAX: (212) 953-7733

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:

LENGTH: 4368 nucleic acids
TYPE: nucleic acid

STRANDEDNESS: double strand
TOPOLOGY: linear

MOLECULE TYPE: cDNA
ORIGINAL SOURCE:

ORGANISM: mouse
TISSUE TYPE: brain

PUBLICATION INFORMATION:

;; AUTHORS: Masayoshi MISHINA
;; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
;; RELEVANT RESIDUES IN SEQ ID NO: 17: FROM 1 TO 4368
US-08-026-138E-17

Query Match 3.2%; Score 24; DB 1; Length 4368;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
DB 597 GCTCTGCAGCTTCTTCAGCTGATT 574

RESULT 2
US-08-026-138E-6/C
; Sequence 6, Application US/08026138E
; Patent No. 5502166
; GENERAL INFORMATION:
; APPLICANT: MASAYOSHI MISHINA
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nishiohata Residence 1-107
; STREET: 5214, Nishiohata-machi
; CITY: Niigata-shi
; STATE: Niigata-ken
; COUNTRY: JAPAN
; ZIP: 951

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS v.5
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/026.138E
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 39563/1992
; FILING DATE: 26-FEB-1992
; APPLICATION NUMBER: JP 173155/1992
; FILING DATE: 30-JUN-1992
; APPLICATION NUMBER: JP 215017/1992
; FILING DATE: 12-AUG-1992
; APPLICATION NUMBER: JP 303878/1992
; FILING DATE: 13-NOV-1992

; ATTORNEY/AGENT INFORMATION:
; NAME: Hamburg, C. Bruce
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-4551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4446 nucleic acids
; TYPE: nucleic acid
; STRANDEDNESS: double strand
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: mouse
; TISSUE TYPE: brain
; PUBLICATION INFORMATION:
; AUTHORS: Masayoshi MISHINA
; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 4446
US-08-026-138E-6

Query Match 3.2%; Score 24; DB 1; Length 4446;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 GCTCTGCAGCTTCTTCAGCTGATT 310
DB 675 GCTCTGCAGCTTCTTCAGCTGATT 652

RESULT 3
US-08-139-862-9
; Sequence 9, Application US/08139862
; Patent No. 5723117
; GENERAL INFORMATION:
; APPLICANT: NAKAI, Satoru
; APPLICANT: AKAMATSU, Sei-ji
; APPLICANT: MASUI, Yoshihiro
; APPLICANT: NISHIDA, Tsutomu
; APPLICANT: KAMOGASHIRA, Takashi
; APPLICANT: HIRAI, Yoshikatu
; TITLE OF INVENTION: AGENT FOR PREVENTING AND TREATING
; TITLE OF INVENTION: HEPATITIS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,862
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/839,770
; FILING DATE: 09-APR-1992
; CLASSIFICATION: 424
; APPLICATION NUMBER: JP 212941/1990
; FILING DATE: 10-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01067
; FILING DATE: 09-AUG-1991

; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-28983
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: synthetic DNA
US-08-139-862-9

Query Match 2.5%; Score 19; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 AATATTATGAGAAATCATCA 371
DB 32 AATATTATGAGAAATCATCA 50

RESULT 4
US-08-139-862-10/c

; Sequence 10, Application US/08139862
; Patent No. 5723117
; GENERAL INFORMATION:
; APPLICANT: NAKAI, Satoru
; APPLICANT: AKAMATSU, Seiji
; APPLICANT: MASUI, Yoshihiro
; APPLICANT: NISHIDA, Tsutomu
; APPLICANT: KAMOGASHIRA, Takashi
; APPLICANT: HIRAI, Yoshikatsu
; TITLE OF INVENTION: AGENT FOR PREVENTING AND TREATING
; TITLE OF INVENTION: HEPATITIS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUCHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,862
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/839,770
; FILING DATE: 09-APR-1992
; CLASSIFICATION: 424
; APPLICATION NUMBER: JP 212941/1990
; FILING DATE: 10-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01067
; FILING DATE: 09-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-28983
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: synthetic DNA
; US-08-139-862-10

Query Match 2.5%; Score 19; DB 1; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 AATATTATGCAATCATCA 371
DB 29 AATATTATGCAATCATCA 11

RESULT 5
US-08-858-207A-205/G
; Sequence 205, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard

; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 205:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-858-207A-205

Query Match 2.5%; Score 19; DB 4; Length 813;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGCTGATTTA 312
DB 634 AGCTTCTTCAGCTGATTTA 616

RESULT 6
US-09-134-078-4/C
; Sequence 4, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 1365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026

```
;
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Haile, Lisa A.
;   REGISTRATION NUMBER: 38,347
;   REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858/677-1456
;   TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1530 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
; FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 1...1527
; US-09-134-078-4
;
; Query Match          2.5%; Score 19; DB 4; Length 1530;
; Best Local Similarity 100.0%; Pred. No. 4.3;
; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 666 AATCTTCGTACATGGGTC 684
      |||||||
DB 1305 AATCTTCGTACATGGGTC 1287
      |||||||

RESULT 7
US-08-920-812-22
; Sequence 22, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6914 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   ORIGINAL SOURCE:
;     ORGANISM: Enterobacter cloacae
;     STRAIN: Clinical Isolate ET-49
; US-08-920-827-22
; Query Match          2.5%; Score 19; DB 1; Length 6914;
; Best Local Similarity 100.0%; Pred. No. 4.1;
; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 190 GAGCTGAGATGTGAACGAC 208
      |||||||
DB 3378 GAGCTGAGATGTGAACGAC 3396
      |||||||

RESULT 8
US-08-920-827-22
; Sequence 22, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6914 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   ORIGINAL SOURCE:
;     ORGANISM: Enterobacter cloacae
;     STRAIN: Clinical Isolate ET-49
; US-08-920-827-22
; Query Match          2.5%; Score 19; DB 1; Length 6914;
; Best Local Similarity 100.0%; Pred. No. 4.1;
; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 190 GAGCTGAGATGTGAACGAC 208
      |||||||
DB 3378 GAGCTGAGATGTGAACGAC 3396
      |||||||
```

```
Db 3378 GAGCTGAGATGTGAACCAG 3396
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6914 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterobacter cloacae
; STRAIN: Clinical isolate ET-49
; US-08-921-177-22

Query Match 2.5%; Score 19; DB 1; Length 6914;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 GAGCTGAGATGTGAACCAG 208
|||||
Db 3378 GAGCTGAGATGTGAACCAG 3396

RESULT 11
US-08-920-828-22
; Sequence 22, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 435
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25

Db 3378 GAGCTGAGATGTGAACCAG 3396
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 435
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
```

```

; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6914 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterobacter cloacae
; STRAIN: Clinical Isolate ET-49
; US-08-920-828-22

Query Match 2.5%; Score 19; DB 2; Length 6914;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 GAGCTGAGATGTGAACCCAG 208
Db 3378 GAGCTGAGATGTGAACCCAG 3396

RESULT 12
US-08-976-259-3
; Sequence 3, Application US/08976259
; Patent No. 6316509
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316509
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CSM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6914 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Enterobacter cloacae
; STRAIN: Clinical Isolate ET-49
; US-08-920-828-22

Query Match 2.5%; Score 19; DB 4; Length 8752;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 TTAATTAACACTTATTATAA 625
Db 756 TTAATTAACACTTATTATAA 774

RESULT 13
US-08-961-527-65/c
; Sequence 65, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-65

Query Match 2.5%; Score 19; DB 4; Length 11831;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGTCGATTTA 312
Db 9464 AGCTTCTTCAGTCGATTTA 9446

RESULT 14
US-09-131-648-3
; Sequence 3, Application US/09131648
; Patent No. 6168920
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
```


FILE REFERENCE: PF-0576 US
CURRENT APPLICATION NUMBER: US/09/131.648
CURRENT FILING DATE: 1998-08-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PERL Program
SEQ ID NO 3
LENGTH: 1643
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2635136
US-09-131-648-3

Query Match 2.4%; Score 18; DB 4; Length 1643;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 ATTTGGTGCAATAAAT 562
|||||
Db 1621 ATTTGGTGCAATAAAT 1638

RESULT 15

US-08-028-463-14/c
Sequence 14, Application US/08028463
Patent No. 5731176

GENERAL INFORMATION:

APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOUCHI, SUCHAYU
APPLICANT: NISHIYAMA, MAKATO
TITLE OF INVENTION: DNA FRAGMENT ENCODING A POLYPEPTIDE
TITLE OF INVENTION: HAVING NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/028.463
FILING DATE: 09-MAR-1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/694,747
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7005-023-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-9741

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 1970 base pairs

TYPE: nucleic acid

STANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Rhodococcus rhodochrous

STRAIN: J-1 (FERM BP-1478)

FEATURE:
NAME/KEY: CDS
LOCATION: 408..1094
FEATURE:
NAME/KEY: CDS
LOCATION: 1111..1719
US-08-028-463-14

Query Match 2.4%; Score 18; DB 1; Length 1970;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 TTCTCATTCCTTTTCATC 243

|||||

Db 407 TTCTCATTCCTTTTCATC 390

Search completed: April 12, 2003, 09:02:50
Job time : 52.8169 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 08:22:00 ; Search time 48.0208 Seconds
(without alignments)
13699.805 Million cell updates/sec

Title: US-09-914-152-3_COPY_1_750

Perfect score: 750

Sequence: 1 cyctctggcaaggtagacc.....ctagaatagaatgatttgt 750

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	21	2.8	831	9	US-09-738-626-1012
2	19	2.5	436	10	US-09-360-352-2104
3	19	2.5	441	10	US-09-815-242-9119
4	19	2.5	441	10	US-09-815-242-9119
5	19	2.5	1530	9	US-10-121-032-4
6	19	2.5	2000	9	US-09-938-842A-4586
7	19	2.5	8752	10	US-09-956-004-3
8	19	2.5	1691139	9	US-10-067-514-1
9	18	2.4	173	9	US-09-738-444-231
10	18	2.4	245	10	US-09-728-446-867
11	18	2.4	378	10	US-09-878-574-3659
12	18	2.4	1643	10	US-09-747-804-3
13	18	2.4	1792	10	US-09-925-300-219
14	18	2.4	1902	10	US-09-349-385-2
15	18	2.4	3477	10	US-09-969-347-221
16	18	2.4	3645	9	US-09-292-758-85
17	18	2.4	5023	9	US-10-078-854-336
18	18	2.4	5023	10	US-09-764-878-336
19	18	2.4	7869	10	US-09-349-385-3

c 20	18	2.4	35414	10	US-09-836-607-45	Sequence 45, Appl
c 21	18	2.4	302250	10	US-09-962-832-154	Sequence 154, Appl
c 22	17	2.3	38	9	US-09-825-805-1035	Sequence 1035, Ap
c 23	17	2.3	68	10	US-09-878-574-7731	Sequence 7731, Ap
c 24	17	2.3	436	10	US-09-867-701-7108	Sequence 7108, Ap
c 25	17	2.3	553	10	US-09-864-761-8586	Sequence 8586, Ap
c 26	17	2.3	1181	9	US-09-938-842A-5224	Sequence 5224, Ap
c 27	17	2.3	1200	10	US-09-887-576-752	Sequence 752, Appl
c 28	17	2.3	2002	10	US-09-887-576-51	Sequence 51, Appl
c 29	17	2.3	2535	10	US-09-815-242-3917	Sequence 3917, Ap
c 30	17	2.3	2538	10	US-09-815-242-6751	Sequence 6751, Ap
c 31	17	2.3	6452	10	US-09-954-456-308	Sequence 308, Appl
c 32	17	2.3	11871	10	US-09-070-927A-171	Sequence 171, Appl
c 33	17	2.3	32193	9	US-09-764-868-1508	Sequence 1508, Ap
c 34	17	2.3	56516	9	US-09-853-526-1	Sequence 1, Appli
c 35	17	2.3	56516	10	US-09-901-484A-1	Sequence 179, Appl
c 36	17	2.3	56520	9	US-09-853-526-179	Sequence 179, Appl
c 37	17	2.3	56520	10	US-09-901-484A-179	Sequence 3, Appli
c 38	17	2.3	98865	10	US-09-770-689A-3	Sequence 17, Appl
c 39	17	2.3	152331	9	US-10-095-407-16	Sequence 16, Appl
c 40	17	2.3	176373	9	US-10-095-407-17	Sequence 3, Appli
c 41	17	2.3	186957	9	US-10-385-770-3	Sequence 4, Appli
c 42	17	2.3	513509	9	US-09-754-853A-4	Sequence 1, Appli
c 43	17	2.3	640681	10	US-09-790-988-1	Sequence 2825, Ap
c 44	16	2.1	188	10	US-09-815-242-2825	Sequence 217, Appl
c 45	16	2.1	191	10	US-09-770-696-217	

ALIGNMENTS

RESULT 1

US-09-738-626-1012
; Sequence 1012, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIKOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1012
; LENGTH: 831
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1012

Query Match 2.8%; Score 21; DB 9; Length 831;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 ATATTGATTTGGCAATTAGTG 351

DB 404 ATATTGATTTGGCAATTAGTG 424

RESULT 2

US-09-960-352-2104
; Sequence 2104, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2104
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 10-BOVMS1-016-Q1-E1-C5
US-09-960-352-2104

Query Match 2.5%; Score 19; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 TTTATTTTAAATAAACTT 618
|||||
DB 67 TTTATTTTAAATAAACTT 85

RESULT 3

US-09-815-242-9119/c
; Sequence 9119, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9119
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(441)
US-09-815-242-9119

RESULT 4

US-09-815-242-9477/c
; Sequence 9477, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9477
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(441)
US-09-815-242-9477

Query Match 2.5%; Score 19; DB 10; Length 441;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGCTGATTTA 312
|||||
DB 405 AGCTTCTTCAGCTGATTTA 387

RESULT 5

US-10-121-032-4/c
; Sequence 4, Application US/10121032
; Patent No. US20020155550A1
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
US-10-121-032-4/c

Query Match 2.5%; Score 19; DB 10; Length 441;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 AGCTTCTTCAGCTGATTTA 312
|||||
DB 405 AGCTTCTTCAGCTGATTTA 387

```

; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: 09-Apr-2002
; FILING DATE: 09-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1527
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-121-032-4
Query Match 2.5%; Score 19; DB 9; Length 1530;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 666 AATCTTCGCTACATGGCTC 684
DB 1305 AATCTTCGCTACATGGCTC 1287
RESULT 6
US-09-938-842A-4586/c
; Sequence 4586, Application US/05938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4586
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4586
Query Match 2.5%; Score 19; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 GATTATTATTTAATTAAAC 616
DB 1743 GATTATTATTTAATTAAAC 1725
RESULT 7
US-09-956-004-3
; Sequence 3, Application US/09956004
; Patent No. US20020072595A1
; GENERAL INFORMATION:
; APPLICANT: Patrick J. Dillon et al.
; TITLE OF INVENTION: Nucleotide Sequences of Escherichia coli Pathogenicity Island
; FILE REFERENCE: PB324D1
; CURRENT APPLICATION NUMBER: US/09/956,004
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: 08/976,259
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/061,953
; PRIOR FILING DATE: 1997-10-14
; PRIOR APPLICATION NUMBER: 60/031,626
; PRIOR FILING DATE: 1996-11-22
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 8752
; TYPE: DNA
; ORGANISM: Escherichia coli
; NAME/KEY: misc.feature
; LOCATION: (16)..(16)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (37)..(37)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (119)..(119)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (2309)..(2309)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (3498)..(3498)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (3645)..(3645)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (6614)..(6614)
; OTHER INFORMATION: n equals a, t, g, or c
US-09-956-004-3
Query Match 2.5%; Score 19; DB 10; Length 8752;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 607 TTAATTAACCTATTATAA 625
DB 756 TTAATTAACCTATTATAA 774
RESULT 8
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
```

```
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match          2.5%; Score 19; DB 9; Length 1691139;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 CTCTGATTCCAAATTCCT 230
      |||||
Db 457119 CTCTGATTCCAAATTCCT 457137

RESULT 9
US-09-728-444-231/c
; Sequence 231, Application US/09728444
; Patent No. US20020161207A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020161207A1e1 Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0100-USA
; CURRENT APPLICATION NUMBER: US/09/728,444
; CURRENT FILING DATE: 2000-11-30
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1206
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231
; LENGTH: 173
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-444-231

Query Match          2.4%; Score 18; DB 9; Length 173;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 589 TTTTTCAGAGATTATT 606
      |||||
Db 167 TTTTTCAGAGATTATT 150

RESULT 10
US-09-728-446-867/c
; Sequence 867, Application US/09728446
; Patent No. US20020081668A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020081668A1e1 Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0101-USA
; CURRENT APPLICATION NUMBER: US/09/728,446
; CURRENT FILING DATE: 2000-11-30
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1461
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 867
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Mus musculus
; NAME/KEY: misc_feature
; LOCATION: (1)...(245)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-446-867

Query Match          2.4%; Score 18; DB 10; Length 245;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 589 TTTTTCAGAGATTATT 606
      |||||
Db 239 TTTTTCAGAGATTATT 222

RESULT 11
US-09-878-574-3659/c
; Sequence 3659, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3659
; LENGTH: 378
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-007-Q1-B1-E3
US-09-878-574-3659

Query Match          2.4%; Score 18; DB 10; Length 378;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 GCAGCTGCAGAACTTCA 452
      |||||
Db 31 GCAGCTGCAGAACTTCA 14

RESULT 12
US-09-747-804-3
; Sequence 3, Application US/09747804
; Patent No. US20010010913A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
; FILE REFERENCE: PF-0576 US
; CURRENT APPLICATION NUMBER: US/09/747,804
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/131,648
; PRIOR FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 1643
; TYPE: DNA
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:52:34 ; Search time 1358.57 Seconds
(without alignments)
16087.603 Million cell updates/sec

Title: US-09-914-152-3_COPY_2000_2750

Perfect score: 751

Sequence: 1 cctgccttggtcccaaaag.....ttagtgaggtggaagatgcc 751

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb.pa.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	751	100.0	10562	6	E38420	Novel polyp
2	700	93.2	170121	9	AF064860	Homo sapi
3	700	93.2	185982	2	AC073231	Homo sapi
4	700	93.2	340000	9	HS21C080	Homo sapi
5	51	6.8	172813	9	AC020550	Homo sapi
6	51	6.8	184931	2	AC114800	Homo sapi
7	50	6.7	117753	9	AC012500	Homo sapi
8	50	6.7	146437	2	AC022836	Homo sapi
9	50	6.7	160558	2	AC025832	Homo sapi
10	50	6.7	163764	9	HSBA64P14	Human DNA
11	50	6.7	169017	2	AL603742	Homo sapi
12	50	6.7	231200	2	AC113399	Homo sapi
13	49	6.5	2418	9	HS049973	Human Tigge
14	49	6.5	92872	2	AC022425	Homo sapi
15	49	6.5	168231	9	AL589863	Human DNA
16	49	6.5	171468	9	AC007394	Homo sapi
17	49	6.5	177518	9	AC093577	Homo sapi
18	49	6.5	183514	2	AC012412	Homo sapi
19	49	6.5	190325	9	AL732327	Human DNA
20	49	6.5	203530	9	AC025097	Homo sapi
21	49	6.5	204024	2	AC109925	Homo sapi
22	48	6.4	56325	9	AL359712	Human DNA
23	48	6.4	187127	9	AC097359	Homo sapi
24	48	6.4	200000	9	AP000493	Homo sapi
25	48	6.4	224574	2	AC069153	Homo sapi
26	47	6.3	56919	2	AC100761	Homo sapi
27	47	6.3	158395	9	AC079600	Homo sapi
28	47	6.3	17262	2	AC123979	Papio cyn
29	47	6.3	183204	2	AC018906	Homo sapi
30	46	6.1	182547	2	AC112696	Homo sapi
31	46	6.1	190956	2	CNS01RHS	Homo sapi
32	46	6.1	200807	9	AC073278	Homo sapi
33	46	6.1	213353	2	AC010258	Homo sapi
34	45	6.0	1725	9	AK094217	Homo sapi
35	45	6.0	44530	9	AC074274	Homo sapi
36	45	6.0	95391	9	AC005926	Homo sapi
37	45	6.0	112123	9	AC095053	Homo sapi
38	45	6.0	137716	9	AC104068	Homo sapi
39	45	6.0	138890	9	AC007970	Homo sapi
40	45	6.0	139111	2	AC114281	Homo sapi
41	45	6.0	139919	9	AC002433	Homo sapi
42	45	6.0	151961	2	AC037443	Homo sapi
43	45	6.0	154716	2	AC069502	Homo sapi
44	45	6.0	157678	9	AC084222	Homo sapi
45	45	6.0	173547	2	AC073637	Homo sapi

ALIGNMENTS

RESULT 1

E38420

LOCUS

DEFINITION Novel polypeptide.

ACCESSION E38420

VERSION E38420.1 GI:18626994

KEYWORDS JP 2000245464-A/2.

SOURCE Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 10562)

AUTHORS Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.

TITLE Novel polypeptide

JOURNAL

Patent: JP 2000245464-A 2 12-SEP-2000;

E38420 10562 bp DNA linear PAT 31-JAN-2002

```
COMMENT
KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464-A/2
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
C12P21/02
PC C12P21/08, C12Q1/68, G01N33/53, (C12N1/21, C12R1:185), (C12N5/10,
C12R1:91),
PC (C12P21/02, C12R1:185), (C12P21/02, C12R1:91), C12N15/00, C12N5/00,
PC (C12N5/00, C12R1:91)
CC
Key Location/Qualifiers
FT promoter (1), (5000)
FT exon (5001), (5140)
FT exon (5011), (5273)
FT exon (5459), (5567)
FT exon (7427), (7586)
FT exon (8234), (10562).
FEATURES
source
1..10562
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2610 a 2415 c 2574 g 2963 t
ORIGIN
Query Match 100.0%; Score 751; DB 6; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTGCCCTTGGCTCCCAAGTCTAGGATTACAGGATGAGCCACCATGTCGACTTCAT 60
DB 2000 CTTGCCCTTGGCTCCCAAGTCTAGGATTACAGGATGAGCCACCATGTCGACTTCAT 2059
QY 61 GATAAACTTCAGTGATGAGGAGTGCCTCTATGATGACAAAGAGTGGTTCTTG 120
DB 2060 GATAAACTTCAGTGATGAGGAGTGCCTCTATGATGACAAAGAGTGGTTCTTG 2119
QY 121 AATGGAATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAAAGAAAT 180
DB 2120 AATGGAATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAAAGAAAT 2179
QY 181 TACAGTGTACATAGAGTTAGTTGATGAAGCAGTAGCAGGATTCGAGAGGATCGATCCCA 240
DB 2180 TACAGTGTACATAGAGTTAGTTGATGAAGCAGTAGCAGGATTCGAGAGGATCGATCCCA 2239
QY 241 ATTCAAATAAGTCTCTGTGGTAAATGCTATCAATGCGGTGCGATGCTACAGAG 300
DB 2240 ATTCAAATAAGTCTCTGTGGTAAATGCTATCAATGCGGTGCGATGCTACAGAG 2299
QY 301 AATCTATCATGAAGGAGAGTCAATGATGTCGCAACTCATCTGTTGCTATTGTTA 360
DB 2300 AATCTATCATGAAGGAGAGTCAATGATGTCGCAACTCATCTGTTGCTATTGTTA 2359
QY 361 AGAATTTGTCAGGACACCCCACTTCAACCAACCATGACCTGATCAGTCAGGAGCAT 420
DB 2360 AGAATTTGTCAGGACACCCCACTTCAACCAACCATGACCTGATCAGTCAGGAGCAT 2419
QY 421 CCACATGAGGAGAGCTCCAGAGTAAAGAGTATGATTTCTTAAGGATCAGATG 480
DB 2420 CCACATGAGGAGAGCTCCAGAGTAAAGAGTATGATTTCTTAAGGATCAGATG 2479
QY 481 AACATTAGCATTTTTTAAGCAATAAGTATTTTACGTAAGATATGATGTTATTTTT 540
DB 2480 AACATTAGCATTTTTTAAGCAATAAGTATTTTACGTAAGATATGATGTTATTTTT 2339
QY 541 AGGCATATGCTATATGCAATTAATGACATCCAGTATATGTAACATCACTTAAATG 600
DB 2540 AGGCATATGCTATATGCAATTAATGACATCCAGTATATGTAACATCACTTAAATG 2599
```

```
QY 601 CACTGGGAGATAAAGATATTGCTCTTTATGATATTGCTTTATGCACTAGTCTGTAA 560
DB 2600 CACTGGGAGATAAAGATATTGCTCTTTATGATATTGCTTTATGCACTAGTCTGTAA 2659
QY 661 TGGAACTACATATTATCTCTTGGGTACACCTGTATACAGAAAGAAATTTATCATGAGAAA 720
DB 2660 TGGAACTACATATTATCTCTTGGGTACACCTGTATACAGAAAGAAATTTATCATGAGAAA 2719
QY 721 TGCTCATGAATGATGAGGCTGGAAGTCC 751
DB 2720 TGCTCATGAATGATGAGGCTGGAAGTCC 2750

RESULT 2
AF064860 170121 bp DNA linear PRI 05-MAR-2002
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
sequence.
AF064860
ACCESSION AF064860.2 GI:18958624
VERSION HTG; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170121)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Horinaka,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagand,E., Raaf,J., Wehrmeyer,S.,
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
10830953
REFERENCE 2 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Feb 27, 2002 this sequence version replaced gi:3171153.
Location/Qualifiers
1..170121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 70124"
BASE COUNT 45184 a 36756 c 37509 g 50672 t
ORIGIN
```


Query Match	93.2%; Score 700; DB 9; Length 170121;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 750; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY 1	CTGCGCTTGGCTCCCAAGTCTAGGATTACAGGATGATGACCCACCATGTCGACTTCAT 60	
DB 83814	CTGCGCTTGGCTCCCAAGTCTAGGATTACAGGATGATGACCCACCATGTCGACTTCAT 83873	
QY 61	GATAAATCTCAGTGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 120	
DB 83874	GATAAATCTCAGTGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 83933	
QY 121	AAATGAATCTACTCTCTGAGTGGTGTGAACATTTGTAATGACAAAGAAAT 180	
DB 83934	AAATGAATCTACTCTCTGAGTGGTGTGAACATTTGTAATGACAAAGAAAT 83993	
QY 181	TACAGTGTACATPAGATGATGATGAGCAGTACGAGGATTCGAGAGGATTCCTCA 240	
DB 83994	TACAGTGTACATPAGATGATGATGAGCAGTACGAGGATTCGAGAGGATTCCTCA 84053	
QY 241	ATTTCATATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 300	
DB 84054	ATTTCATATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84113	
QY 301	AAATCTATCATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 360	
DB 84114	AAATCTATCATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84173	
QY 361	AGAAATGTGACGAGCAGCAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 420	
DB 84174	AGAAATGTGACGAGCAGCAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84233	
QY 421	CCACATGAGGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 480	
DB 84234	CCACATGAGGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84293	
QY 481	ACATAGGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 540	
DB 84294	ACATAGGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84353	
QY 541	AGGATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 600	
DB 84354	AGGATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84413	
QY 601	CATGAGGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 660	
DB 84414	CATGAGGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84473	
QY 661	TGGAATCATATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 720	
DB 84474	TGGAATCATATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84533	
QY 721	TGCTCATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 751	
DB 84534	TGCTCATGATGATGAGGAGTGGCTCTTATGATGACAAAGAGTGTTCTTG 84564	
RESULT 3		
AC073231	185982 bp DNA linear HTG 17-JUN-2000	
LOCUS		
DEFINITION	Homo sapiens chromosome 17 clone RP11-729A4, WORKING DRAFT	
SEQUENCE	SEQUENCE, 25 unordered pieces.	
AC073231		
AC073231.1	GI:8440043	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	Homo sapiens	
ORGANISM	Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
Waterston, R.H.		
1 (bases 1 to 185982)		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
2 (bases 1 to 185982)		
Waterston, R.H.		
Direct Submission		
Submitted (10-JUN-2000)		
Genome Sequencing Center, Washington		
University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
MO 63108, USA		
NOTE: This is a 'working draft' sequence. It currently		
consists of 25 contigs. The true order of the pieces		
is not known and their order in this sequence record is		
arbitrary. Gaps between the contigs are represented as		
runs of N, but the exact sizes of the gaps are unknown.		
This record will be updated with the finished sequence		
as soon as it is available and the accession number will		
be preserved.		
1724: contig of 1724 bp in length		
1725: gap of unknown length		
1825: contig of 1147 bp in length		
2972: gap of unknown length		
3072: contig of 1589 bp in length		
4661: gap of unknown length		
5930: contig of 1170 bp in length		
6030: gap of unknown length		
8127: contig of 2097 bp in length		
8227: gap of unknown length		
8228: contig of 1933 bp in length		
10160: gap of unknown length		
10261: contig of 2976 bp in length		
13236: gap of unknown length		
13237: contig of 3267 bp in length		
16604: gap of unknown length		
20684: contig of 3981 bp in length		
20785: gap of unknown length		
24835: contig of 4051 bp in length		
24935: gap of unknown length		
29727: contig of 4792 bp in length		
29827: gap of unknown length		
34947: contig of 5120 bp in length		
35047: gap of unknown length		
41164: contig of 6117 bp in length		
41265: gap of unknown length		
47050: contig of 5786 bp in length		
47150: gap of unknown length		
53084: contig of 5934 bp in length		
53184: gap of unknown length		
60624: contig of 7440 bp in length		
60724: gap of unknown length		
69057: contig of 8333 bp in length		
69157: gap of unknown length		
78529: contig of 9172 bp in length		
78629: gap of unknown length		
85475: contig of 6846 bp in length		
85575: gap of unknown length		
97151: contig of 11576 bp in length		
97251: gap of unknown length		
113111: contig of 15860 bp in length		
113211: gap of unknown length		
129870: contig of 16659 bp in length		
129871: gap of unknown length		
145568: contig of 15598 bp in length		
145569: gap of unknown length		
164596: contig of 18928 bp in length		
164597: gap of unknown length		
185982: contig of 21286 bp in length.		
Location/Qualifiers		
1. 185982		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/chromosome="17"		
/clone="RP11-729A4"		
1. 1724		
/note="assembly_name:Contig3"		
misc_feature		
1. 1724		

```
misc_feature      1825..2971
/note="assembly_name:Contig4"
misc_feature      3072..4660
/note="assembly_name:Contig5"
misc_feature      4761..5330
/note="assembly_name:Contig6"
vector_side:right
clone_end:T7
6031..8127
/note="assembly_name:Contig7"
8228..10160
/note="assembly_name:Contig8"
10261..13236
/note="assembly_name:Contig9"
13337..16603
/note="assembly_name:Contig10"
16704..20684
/note="assembly_name:Contig11"
20785..24835
/note="assembly_name:Contig12"
24936..29727
/note="assembly_name:Contig13"
clone_end:SP6
vector_side:right
29828..34947
/note="assembly_name:Contig14"
35048..41164
/note="assembly_name:Contig15"
41265..47050
/note="assembly_name:Contig16"
47151..53084
/note="assembly_name:Contig17"
53185..60624
/note="assembly_name:Contig18"
60725..69057
/note="assembly_name:Contig19"
69158..78529
/note="assembly_name:Contig20"
78630..85475
/note="assembly_name:Contig21"
85576..97151
/note="assembly_name:Contig22"
97252..113111
/note="assembly_name:Contig23"
113212..123870
/note="assembly_name:Contig24"
129971..145568
/note="assembly_name:Contig25"
145669..164596
/note="assembly_name:Contig26"
164697..185982
/note="assembly_name:Contig27"
BASE COUNT 53704 a 39120 c 39561 g 51192 t 2405 others
ORIGIN
Query Match      93.2%; Score 700; DB 2; Length 185982;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTTGCCCTTGGCTCCCAAAAGTCTAGATTACAGGATGAGCCACCATGGTCGACTTCAT 60
|||||
Db 26697 CTTGCCCTTGGCTCCCAAAAGTCTAGATTACAGGATGAGCCACCATGGTCGACTTCAT 26756
|||||
QY 61 GATAAACTTCAGTGGATGAGGAGTGCCTCTTATCATGAACAAAGAGGTGTTCTTG 120
|||||
Db 26757 GATAAACTTCAGTGGATGAGGAGTGCCTCTTATCATGAACAAAGAGGTGTTCTTG 26816
|||||
QY 121 AAATGAATCTACTCTGCTGGAGAGCTGTGAACATGTTGAATGACAGAAGAATT 180
|||||
Db 26817 AAATGAATCTACTCTGCTGGAGAGTGTGTGAACATGTTGAATGACAGAAGAATT 26876
|||||
QY 181 TACAGTGTACATAGAGTTAGTTGATGAAGCAGTAGTACAGGATTCGAGAGGATCGATTCCA 240
|||||
```

```
Db 26877 TACAGTGTACATAGAGTTAGTTGATGAAGCAGTAGTACAGGATTCGAGAGGATCGATTCCA 26936
QY 241 ATTCAAATAAAGTCTCTCTGGGTAAATAGTATCAAAATGGCGTGGCGATCGTACAGAG 300
|||||
Db 26937 ATTCAAATAAAGTCTCTCTGGGTAAATAGTATCAAAATGGCGTGGCGATCGTACAGAG 26996
|||||
QY 301 AAATCTATCATGAAGAAGAGTCAATTGATGTGGCAAACTTCATTGTTGTCGTATTTTA 360
|||||
Db 26997 AAATCTATCATGAAGAAGAGTCAATTGATGTGGCAAACTTCATTGTTGTCGTATTTTA 27056
|||||
QY 361 AGAATTGTGAGGACCCACCCCAACCTTCAACAACCATGACCCCTGATCGTACGAGGCCAT 420
|||||
Db 27057 AGAATTGTGAGGACCCACCCCAACCTTCAACAACCATGACCCCTGATCGTACGAGGCCAT 27116
|||||
QY 421 CCACATTGAGCGGAGAACCTCCAGCAGTAAAGATTATGATCTCTAAAGGATCAGATG 480
|||||
Db 27117 CCACATTGAGCGGAGAACCTCCAGCAGTAAAGATTATGATCTCTAAAGGATCAGATG 27176
|||||
QY 481 AACATTAGCATTTTTTTAAGCAATAAAAGTATTTTTACGTAAGATATGTTATGTTTATTTT 540
|||||
Db 27177 AACATTAGCATTTTTTTAAGCAATAAAAGTATTTTTACGTAAGATATGTTATGTTTATTTT 27236
|||||
QY 541 AGGCATATGCTATTATGTCATTAAATAGACTCCAGTATATTGTAACATAAATTTAAATG 600
|||||
Db 27237 AGGCATATGCTATTATGTCATTAAATAGACTCCAGTATATTGTAACATAAATTTAAATG 27296
|||||
QY 601 CACTGGGAGATAAAAGTATTGCTCTTTTATGATATTGTTGCTTTATTGTCAGTAGTCTGTAA 660
|||||
Db 27297 CACTGGGAGATAAAAGTATTGCTCTTTTATGATATTGTTGCTTTATTGTCAGTAGTCTGTAA 27356
|||||
QY 661 TGGAACTACATATCTCTTGGGTACACCTGTATACAGAAGAAATTTATCATGAGGAAA 720
|||||
Db 27357 TGGAACTACATATCTCTTGGGTACACCTGTATACAGAAGAAATTTATCATGAGGAAA 27416
|||||
QY 721 TGCTCATGCAATGATGAGGCTGGAAGTCC 751
|||||
Db 27417 TGCTCATGCAATGATGAGGCTGGAAGTCC 27447
|||||
```

RESULT 4
HS21C080 340000 bp DNA linear PRI 24-MAY-2000
LOCUS Homo sapiens chromosome 21 segment HS21C080.
DEFINITION AL163280 AP001735 BA000005
ACCESSION AL163280.2 GI:7717369
VERSION
KEYWORDS
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)

REFERENCE

AUTHORS

Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
Horisch,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesselmann,L., Dagnid,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.

TITLE

JOURNAL

Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)

COMMENT

The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *


```
repeat_region complement(11315..11556)
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 11702..11753
/notes="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
exon 12649..12678
/genes="SH3BGR"
/number=5
repeat_region complement(13115..13425)
/notes="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(13598..13892)
/notes="Alu"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 14013..14131
/notes="(CA)n"
/rpt_family="simple_repeat"

Query Match 93.2%; Score 700; DB 9; Length 340000;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 750; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTGCTTGGGCTCCAAAGTGTAGGATTACAGGATGAGCCACCATGTCGACTTCAT 60
|||||
Db 157222 CCTGCTTGGGCTCCAAAGTGTAGGATTACAGGATGAGCCACCATGTCGACTTCAT 157281

QY 61 GATAAATCTCAGTGGATGAGGAGTGCCTCTTATGATGACAAAGAGTGTTCTTCTG 120
|||||
Db 157282 GATAAATCTCAGTGGATGAGGAGTGCCTCTTATGATGACAAAGAGTGTTCTTCTG 157341

QY 121 AAATGAATCTACTCCTGCTGAAGTGTGTGAATCTGTGAATGACAAAGAAAT 180
|||||
Db 157342 AAATGAATCTACTCCTGCTGAAGTGTGTGAATCTGTGAATGACAAAGAAAT 157401

QY 181 TACAGTGTACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
|||||
Db 157402 TACAGTGTACATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 157461

QY 241 ATTCAAAATAGTCTTCTGTGGTAAATGCTATCAATGCGCTGCGATGACAGAG 300
|||||
Db 157462 ATTCAAAATAGTCTTCTGTGGTAAATGCTATCAATGCGCTGCGATGACAGAG 157521

QY 301 AAATCTATCATGAAGAGAGTCAATGATGTCGCCAACTTCATGTTCTGCTATTTA 360
|||||
Db 157522 AAATCTATCATGAAGAGAGTCAATGATGTCGCCAACTTCATGTTCTGCTATTTA 157581

QY 361 AGAATTTGTCAGGACCCCAACCTTCAACACCATGACCTGATGATGATGATGATGAT 420
|||||
Db 157582 AGAATTTGTCAGGACCCCAACCTTCAACACCATGACCTGATGATGATGATGATGAT 157641

QY 421 CCACATGAGGCGAGAACCTCCAGCAGTAAAGATGATGATGATGATGATGATGATGAT 480
|||||
Db 157642 CCACATGAGGCGAGAACCTCCAGCAGTAAAGATGATGATGATGATGATGATGAT 157701

QY 481 ACATTTAGCATTTTATGAGCAATAAATTTTACGTAGATGATGATGATGATGATTTT 540
|||||
Db 157702 ACATTTAGCATTTTATGAGCAATAAATTTTACGTAGATGATGATGATGATGATTTT 157761

QY 541 AGCATTAATGCTATTATGATTAATAGACTCCAGTATGATGATGATGATGATGATGAT 600
|||||
Db 157762 AGCATTAATGCTATTATGATTAATAGACTCCAGTATGATGATGATGATGATGATGAT 157821

QY 601 CACTGGAGATAAAGTATTTCTCTTTTATGATATTTTCTTTTATGATGATGATGATGAT 660
|||||
Db 157822 CACTGGAGATAAAGTATTTCTCTTTTATGATATTTCTTTTATGATGATGATGATGAT 157881

QY 661 TGAAGACTACATATCTCTTGGGTACACCTGTATACAGAAAGAAATTTATCATGAGGAAA 720
|||||
```

```
Db 157882 TGAAGACTACATATCTCTTGGGTACACCTGTATACAGAAAGAAATTTATCATGAGGAAA 157941
QY 721 TGCTCATGAATGATGAGGCTGGAAAGTCC 751
|||||
Db 157942 TGCTCATGAATGATGAGGCTGGAAAGTCC 157972
```

```
RESULT 5
AC020550/c 172813 bp DNA linear PRI 09-MAY-2001
LOCUS Homo sapiens BAC clone RP11-198M19 from 2, complete sequence.
DEFINITION AC020550
ACCESSION AC020550
VERSION AC020550.4 GI:10047985
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172813)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 172813)
AUTHORS Du, H., Drone, K., Hawkins, M., Elliott, G. and Waligorski, J.
TITLE The sequence of Homo sapiens BAC clone RP11-198M19
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 172813)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 172813)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 172813)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 10, 2000 this sequence version replaced gi:7630889.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0198M19
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-198M19; actual end is at base position 172813 of RP11-198M19.

FEATURES

Source

1. 172813

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-198M19"

/clone.lib="RPCI-11"

2. 194

/rpt_family="L1"

195. 493

/rpt_family="Alu"

494. 1271

/rpt_family="L1"

1285. 1413

/rpt_family="L1"

1414. 1621

/rpt_family="L1"

1622. 1776

/rpt_family="L1"

1777. 2076

/rpt_family="Alu"

2081. 2378

/rpt_family="Alu"

2162. 2417

/note="similar to EST AW867673 (NID:g8001829)"

2379. 3605

/rpt_family="L1"

3645. 3765

/rpt_family="Alu"

3777. 5106

/rpt_family="L1"

5107. 5336

/rpt_family="Alu"

5345. 5711

/rpt_family="L1"

5712. 5999

/rpt_family="Alu"

6000. 6116

/rpt_family="L1"

6426. 6657

/rpt_family="MIR"

6741. 7033

/rpt_family="Alu"

7067. 7350

/rpt_family="Alu"

7984. 8266

/rpt_family="Alu"

9407. 9533

/rpt_family="Alu"

9558. 9720

/rpt_family="MER1_type"

10027. 10250

/rpt_family="MIR"

10287. 10505

/rpt_family="MIR"

10734. 11025

/rpt_family="Alu"

11557. 11853

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region
11987. 12145
/rpt_family="Alu"
repeat_region
12146. 12463
/rpt_family="Alu"
repeat_region
12464. 12727
/rpt_family="MER1_type?"
repeat_region
12859. 12930
/rpt_family="Mariner"
repeat_region
13149. 13305
/rpt_family="CR1"
repeat_region
13334. 13504
/rpt_family="L2"
repeat_region
13842. 13856
/rpt_family="CR1"
repeat_region
13862. 14167
/rpt_family="MER2_type"
repeat_region
14835. 14923
/rpt_family="L1"
repeat_region
14924. 15232
/rpt_family="Alu"
repeat_region
15233. 16020
/rpt_family="L1"
repeat_region
16203. 16441
/rpt_family="L1"
repeat_region
16472. 16595
/rpt_family="L1"
repeat_region
17579. 17739
/rpt_family="L2"
repeat_region
19061. 19357
/rpt_family="Alu"
misc_feature
19649. 19686
/note="similar to EST AW199183 (NID:g6479752) dal7b12.x1"
repeat_region
19659. 19802
/rpt_family="Alu"
repeat_region
20532. 20826
/rpt_family="Alu"
repeat_region
21118. 21461
/rpt_family="MER2_type"
repeat_region
22523. 22734
/rpt_family="L1"
repeat_region
23178. 23494
/rpt_family="Alu"
repeat_region
23617. 23867
/rpt_family="Alu"
repeat_region
23889. 24186
/rpt_family="L1"
repeat_region
25329. 25417
/rpt_family="MER1_type"
misc_feature
28953. 29131
/note="similar to EST AI541180 (NID:g4458553)"
misc_feature
32701. 32852
/note="similar to EST R13320 (NID:g766396) yf75f04.rl"
repeat_region
33560. 33679
/rpt_family="MIR"
repeat_region
33962. 34339
/rpt_family="MIR"
misc_feature
35350. 35560
/note="similar to EST R13320 (NID:g766396) yf75f04.rl"
repeat_region
36481. 36532
/rpt_family="L2"
repeat_region
36547. 36858
/rpt_family="Alu"

Query Match

Best Local Similarity 6.8%; Score 51; DB 9; Length 172813;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAATGGAATCTACTCCTGGTGAAGATGCTGTGAACATTGTGAAATGACAA 171
|||||

DB 42454 AAATGGAATCTACTCCTGGTGAAGATGCTGTGAACATTGTGAAATGACAA 42404
|||||

```

RESULT 6
AC114800      188491 bp      DNA      linear      HTG 20-JUN-2002
LOCUS      Homo sapiens chromosome 2 clone RP11-776M7, WORKING DRAFT SEQUENCE,
DEFINITION      1 unordered piece.
AC114800
AC114800.4 GI:21327634
VERSION      HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 188491)
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
2 (bases 1 to 188491)
Direct Submission
Submitted (11-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 188491)
Waterston,R.H.
Direct Submission
Submitted (20-JUN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 6, 2002 this sequence version replaced gi:21307583.

REFERENCE
AUTHORS
TITLE
JOURNAL

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0776M07
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 179040 bases at least Q40
Consensus quality: 180936 bases at least Q30
Consensus quality: 188424 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 188491; sum-of-contigs
Quality coverage: 39.67 in Q20 bases; agarose-fp
Quality coverage: 6.31 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 188491: contig of 188491 bp in length.
Location/Qualifiers
1..188491
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-776M7"
1..188491
/note="assembly_name:Contig25
clone_end:SP6
vector_side:right"

```

FEATURES

source

misc_feature

```

54909 a 37185 c 38025 g 58372 t
ORIGIN

```

```

Query Match      6.8%; Score 51; DB 2; Length 188491;
Best Local Similarity 100.0%; Pred. No. 6e-18;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAATGAATCTACTCTCTGCTGAAGATGCTGTGAACATTGTTGAAATGACAA 171
|||||
DB 53218 AAATGAATCTACTCTCTGCTGAAGATGCTGTGAACATTGTTGAAATGACAA 53268
|||||

RESULT 7
AC012500/c
LOCUS      Homo sapiens BAC clone RP11-434M17 from 2, complete sequence.
DEFINITION      AC012500
AC012500
AC012500.7 GI:15144462
VERSION      HTG.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.E. and Waterston,R.
1 (bases 1 to 117763)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED
9847074
2 (bases 1 to 117763)
Isak,A. and Haakenson,W.
The sequence of Homo sapiens BAC clone RP11-434M17
Unpublished
Waterston,R.H.
3 (bases 1 to 117763)
Direct Submission
Submitted (28-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 117763)
Waterston,R.
Direct Submission
Submitted (09-AUG-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 117763)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:13431189.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0434M17

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis

MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-67G7; the clone sequenced to the right is RP11-418N16. Actual start of this clone is at base position 1 of RP11-434M17.

FEATURES

Location/Qualifiers

Source

1. 117763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

/clone="RP11-434M17"

/clone.lib="RP11-11"

1205..1285

/rpt_family="L1"

1286..1321

/rpt_family="(TA)n"

1322..1688

/rpt_family="L1"

1946..2073

/rpt_family="MaLR"

2666..2913

/rpt_family="L1"

3018..3428

/rpt_family="MaLR"

3840..3876

/rpt_family="(CA)n"

4779..4893

/rpt_family="(TA)n"

5348..5485

/rpt_family="L1"

5488..5659

/rpt_family="L1"

5660..5937

/rpt_family="Alu"

5938..6053

/rpt_family="L1"

6075..6278

/rpt_family="L1"

6974..7024

/rpt_family="AT_rich"

7137..7166

/rpt_family="AT_rich"

7514..7724

/rpt_family="L1"

8321..8655

/rpt_family="L1"

8659..8834

/rpt_family="AchoBo"

8838..9577

/rpt_family="L1"

9863..9906

/rpt_family="AT_rich"

10324..10657

/rpt_family="MaLR"

11408..11880

/rpt_family="ERVL"

11930..12366

/rpt_family="ERV1"

12497..12579

/rpt_family="Ct-rich"

repeat_region 12686..13816
/rpt_family="ERVK"
repeat_region 16104..16158
/rpt_family="AT_rich"
repeat_region 16276..17015
/rpt_family="L1"
repeat_region 17043..17226
/rpt_family="L1"
repeat_region 19823..19856
/rpt_family="(CA)n"
repeat_region 21236..21375
/rpt_family="MIR"
repeat_region 22320..22524
/rpt_family="MIR"
repeat_region 22862..23265
/rpt_family="MaLR"
repeat_region 23277..23949
/rpt_family="L2"
repeat_region 24425..24446
/rpt_family="AT_rich"
repeat_region 24594..24731
/rpt_family="MIR"
repeat_region 25382..25455
/rpt_family="MIR"
repeat_region 26504..26525
/rpt_family="AT_rich"
repeat_region 26736..26758
/rpt_family="(TAAAA)n"
repeat_region 27008..27197
/rpt_family="L2"
repeat_region 27226..27729
/rpt_family="L2"
repeat_region 27737..27943
/rpt_family="L2"
repeat_region 28269..28299
/rpt_family="AT_rich"
repeat_region 28826..28854
/rpt_family="AT_rich"
repeat_region 29044..29064
/rpt_family="AT_rich"
repeat_region 29396..29524
/rpt_family="L1"
repeat_region 29990..30658
/rpt_family="MER2_type"
repeat_region 31476..31757
/rpt_family="Alu"
repeat_region 31834..32088
/rpt_family="MER1_type"
repeat_region 32880..32940
/rpt_family="(CATATA)n"
repeat_region 33098..33565
/rpt_family="ERVL"
repeat_region 34052..34358
/rpt_family="ERVL"
repeat_region 34841..35275
/rpt_family="ERVL"
repeat_region 35370..36570
/rpt_family="ERVL"
repeat_region 36732..37291
/rpt_family="ERVL"
repeat_region 37567..38402
/rpt_family="ERVL"
repeat_region 38564..38722
/rpt_family="ERVL"
repeat_region 40310..40599
/rpt_family="Alu"

Query Match

Best Local Similarity 6.7%; Score 50; DB 9; Length 117763;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 CTTGAATGGAATCTACTCTCGTGAAGATGCTGTGAACATTTTGAAT 166

|||||

Db 39315 CTTGAAATGGAATCTACTCTGGTGAAGATGCTGTGTGAACATTGTGAAAT 39266

RESULT 8

AC022836/c

LOCUS

DEFINITION

AC022836

AC022836

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

AC022836 146437 bp DNA linear HTG 12-MAR-2000
Homo sapiens chromosome 11 clone RP11-162I14 map 11, WORKING DRAFT
SEQUENCE, 12 unordered pieces.

AC022836 2 GI:7229854
HIG: HTGS_PHASE1; HTGS_DRAFT.

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 146437)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abramson, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Becker, R., Breda, F.,
Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardina, S., Grant, G., Hago, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Lander, T., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Menees, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A., and Zody, M.

2 (bases 1 to 146437)
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 12, 2000 this sequence version replaced gi:6922254.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6144
Center clone name: L6144

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138795 bases at least Q40
Consensus quality: 142787 bases at least Q30
Consensus quality: 144305 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 145337; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
consists of 12 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 2682: contig of 2682 bp in length

TITLE

JOURNAL

COMMENT

* 2683 2782: gap of 100 bp
2783 5744: contig of 2962 bp in length
5745 5844: gap of 100 bp
5845 12632: contig of 6788 bp in length
12633 12732: gap of 100 bp
12733 20748: contig of 8016 bp in length
20749 20848: gap of 100 bp
20849 30947: contig of 10099 bp in length
30948 31047: gap of 100 bp
31048 40953: contig of 9906 bp in length
40954 41053: gap of 100 bp
41054 51431: contig of 10378 bp in length
51432 51531: gap of 100 bp
51532 63688: contig of 12158 bp in length
63689 63790: gap of 100 bp
63791 77331: contig of 13542 bp in length
77332 77431: gap of 100 bp
77432 93050: contig of 15619 bp in length
93051 93150: gap of 100 bp
93151 116006: contig of 22856 bp in length
116007 116106: gap of 100 bp
116107 146437: contig of 30331 bp in length.

FEATURES

source

1. 146437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"

/clone="RP11-162I14"
/clone_lib="RPC1-11 Human Male BAC"
1. 2682
/note="assembly_fragment"

misc_feature

2783..5744
/note="assembly_fragment"

misc_feature

clone_end:T7
vector_side:left

misc_feature

5845..12632
/note="assembly_fragment"

misc_feature

12733..20748
/note="assembly_fragment"

misc_feature

20849..30947
/note="assembly_fragment"

misc_feature

31048..40953
/note="assembly_fragment"

misc_feature

clone_end:SP6
vector_side:left

misc_feature

41054..51431
/note="assembly_fragment"

misc_feature

51532..63688
/note="assembly_fragment"

misc_feature

63790..77331
/note="assembly_fragment"

misc_feature

77432..93050
/note="assembly_fragment"

misc_feature

93151..116006
/note="assembly_fragment"

misc_feature

116107..146437
/note="assembly_fragment"

BASE COUNT

46075 a 26313 c 26080 g 46869 t
ORIGIN
Query Match 6.7%; Score 50; DB 2: Length 146437;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 CTTGAAATGGAATCTACTCTGGTGAAGATGCTGTGAACATTGTGAAAT 166
|||||
Db 143162 CTTGAAATGGAATCTACTCTGGTGAAGATGCTGTGAACATTGTGAAAT 143113

RESULT 9
AC025832/c
LOCUS

AC025832 AC025832 160558 bp DNA linear HTG 30-MAR-2000


```

DEFINITION Homo sapiens clone RP11-25F19, WORKING DRAFT SEQUENCE, 15 unordered
pieces.
ACCESSION AC025832
VERSION 2
KEYWORDS HTGS, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160558)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-25F19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160558)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
Campolano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galligan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menes, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, K., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teschaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:7249060.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4557
Center clone name: 25_F19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150706 bases at least Q40
Consensus quality: 155861 bases at least Q30
Consensus quality: 157811 bases at least Q20
Insert size: 157000; agarose-fp
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1006: contig of 1006 bp in length
* 1007 1106: gap of 100 bp
* 1107 2305: contig of 1199 bp in length

```

```

* 2206 2405: gap of 100 bp
* 2406 2806: contig of 401 bp in length
* 2807 2906: gap of 100 bp
* 2907 6298: contig of 3392 bp in length
* 6299 6398: gap of 100 bp
* 6399 7953: contig of 1555 bp in length
* 7954 8053: gap of 100 bp
* 8054 13929: contig of 5876 bp in length
* 13930 14029: gap of 100 bp
* 14030 22806: contig of 8777 bp in length
* 22807 22906: gap of 100 bp
* 22907 32351: contig of 9445 bp in length
* 32352 32451: gap of 100 bp
* 32452 41128: contig of 8677 bp in length
* 41129 41228: gap of 100 bp
* 41229 55888: contig of 14560 bp in length
* 55889 55988: gap of 100 bp
* 55989 71584: contig of 15596 bp in length
* 71585 71684: gap of 100 bp
* 71685 90527: contig of 18843 bp in length
* 90528 90627: gap of 100 bp
* 90628 107160: contig of 16533 bp in length
* 107161 107260: gap of 100 bp
* 107261 127617: contig of 20357 bp in length
* 127618 127717: gap of 100 bp
* 127718 160558: contig of 32841 bp in length.
FEATURES
Location/Qualifiers
source
1..160558
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-25F19"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature
1..1006
/note="assembly_fragment"
misc_feature
1107..2305
/note="assembly_fragment"
misc_feature
2406..2806
/note="assembly_fragment"
misc_feature
clone_end:SP6
vector_side:left"
misc_feature
2907..6298
/note="assembly_fragment"
misc_feature
6399..7953
/note="assembly_fragment"
misc_feature
clone_end:R7
vector_side:right"
misc_feature
8054..13929
/note="assembly_fragment"
misc_feature
14030..22806
/note="assembly_fragment"
misc_feature
22907..32351
/note="assembly_fragment"
misc_feature
32452..41128
/note="assembly_fragment"
misc_feature
41229..55888
/note="assembly_fragment"
misc_feature
55989..71584
/note="assembly_fragment"
misc_feature
71685..90527
/note="assembly_fragment"
misc_feature
90628..107160
/note="assembly_fragment"
misc_feature
107261..127617
/note="assembly_fragment"
misc_feature
127718..160558
/note="assembly_fragment"
BASE COUNT 49133 a 29829 c 29905 g 50277 t 1413 others
ORIGIN

```

```

Query Match 6.7%; Score 50; DB 2; Length 160558;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```
repeat_region 28202..28585
/Note="MLT2E repeat: matches 1..397 of consensus"
repeat_region 28586..28705
/Note="MIR repeat: matches 20..140 of consensus"
repeat_region 28621..29099
/Note="MIR repeat: matches 10..262 of consensus"
repeat_region 29431..29566
/Note="MIR repeat: matches 80..226 of consensus"
repeat_region 30015..30215
/Note="MIR repeat: matches 49..257 of consensus"
repeat_region 30475..30927
/Note="MLT1F repeat: matches 100..541 of consensus"
repeat_region 31005..31048
/Note="22 copies 2 mer tt 77% conserved"
repeat_region 31053..34717
/Note="L1M1 repeat: matches 2588..6300 of consensus"
misc_feature 34845..35488
/Note="match: GSS: Em:AQ013989"
misc_feature 34858..35546
/Note="match: GSS: Em:AQ238728"
misc_feature 34893..35261
/Note="match: GSS: Em:B83155"
repeat_region 34929..35059
/Note="L1P2 repeat: matches 6018..6151 of consensus"
repeat_region 35080..36417
/Note="L1P5 repeat: matches 4436..5790 of consensus"
repeat_region 38713..38909
/Note="MIR repeat: matches 19..214 of consensus"
repeat_region 39112..39224
/Note="L2 repeat: matches 2616..2743 of consensus"
repeat_region 39359..39602
/Note="MIR repeat: matches 3..253 of consensus"
repeat_region 41501..41850
/Note="MLT1C repeat: matches 117..466 of consensus"
repeat_region 42411..42526
/Note="MS9A repeat: matches 48..188 of consensus"
repeat_region 42579..42709
/Note="FLAM_A repeat: matches 1..133 of consensus"
repeat_region 42961..43185
/Note="MIR repeat: matches 2..240 of consensus"
repeat_region 43521..43647
/Note="L2 repeat: matches 2585..2710 of consensus"
repeat_region 44192..44459
/Note="MIR repeat: matches 6..261 of consensus"
repeat_region 45666..46121
/Note="L2 repeat: matches 2275..2734 of consensus"
repeat_region 46973..47067
/Note="MIR repeat: matches 1..106 of consensus"
repeat_region 47073..47269
/Note="MER3 repeat: matches 3..209 of consensus"
repeat_region 48802..48899
/Note="L2 repeat: matches 2598..2732 of consensus"
repeat_region 49748..49847
/Note="L2 repeat: matches 2645..2742 of consensus"
repeat_region 49869..50408
/Note="L1P16 repeat: matches 5590..6146 of consensus"
repeat_region 50854..51218
/Note="L2 repeat: matches 2055..2438 of consensus"
repeat_region 51400..51595
/Note="L2 repeat: matches 2553..2750 of consensus"
repeat_region 52015..52087

Query Match 6.7% Score 50; DB 9; Length 163764;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 AATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAGAAAG 176
|||||
Db 67333 AATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAGAAAG 67284

RESULT 11
AL603742/c

LOCUS AL603742 169017 bp DNA linear HTG 11-SEP-2001
DEFINITION Homo sapiens chromosome X clone RP13-295D3, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL603742
VERSION AL603742.3 GI:15591740
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169017)
Direct Submission
Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 12, 2001 this sequence version replaced gi:15209424.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: B829503
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 165021 bases at least Q40
Consensus quality: 166455 bases at least Q30
Consensus quality: 167078 bases at least Q20
Insert size: 169017; sum-of-contigs
Quality coverage: 6.69x in Q20 bases; sum-of-contigs Quality
coverage: 6.79x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1..169017
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-295D3"
/clone_lib="RPCI-13.2"
1..169017
misc_feature 1..169017
/Note="assembly fragment:03121"
BASE COUNT 54965 a 32058 c 31304 g 50690 t
ORIGIN

Query Match 6.7% Score 50; DB 2; Length 169017;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 AATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAGAAAG 176
|||||
Db 106364 AATCTACTCTGTGAAGATGCTGTGAACATTTGTAATGACAGAAAG 106315

RESULT 12
AC113399/c
LOCUS AC113399 231200 bp DNA linear HTG 01-MAR-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-437E23, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC113399
VERSION AC113399.1 GI:19033513
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 231200)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 231200)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 574292
Center clone name: RPCI-11_437E23

Summary Statistics
Consensus quality: 209418 bases at least Q40
Consensus quality: 219818 bases at least Q30
Consensus quality: 224243 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Quality coverage: 9.22 in Q20 bases; agarose-fp estimation
Quality coverage: 7.04 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1357: contig of 1357 bp in length
1358 1457: gap of unknown length
1458 2937: contig of 1540 bp in length
2938 3097: gap of unknown length
3098 4175: contig of 1078 bp in length
4176 4275: gap of unknown length
4276 5565: contig of 1290 bp in length
5566 6998: gap of unknown length
6999 7099: gap of unknown length
7099 8411: contig of 1312 bp in length
8411 8510: gap of unknown length
8510 10115: contig of 1605 bp in length
10116 10215: gap of unknown length
10216 11509: contig of 1294 bp in length
11510 13369: contig of 1760 bp in length
13370 13469: gap of unknown length
13470 14637: contig of 1168 bp in length
14638 14737: gap of unknown length
14738 17408: contig of 2671 bp in length
17409 17508: gap of unknown length
17509 20508: contig of 3000 bp in length
20509 20608: gap of unknown length
20609 25980: contig of 5372 bp in length
25981 26080: gap of unknown length
26081 34402: contig of 8322 bp in length
34403 34502: gap of unknown length
34503 43673: contig of 9171 bp in length
43674 43773: gap of unknown length
43774 54380: contig of 10607 bp in length
54381 54480: gap of unknown length
54481 67192: contig of 12712 bp in length
67193 67292: gap of unknown length
67293 89134: contig of 21842 bp in length
89135 89235: gap of unknown length
89236 110109: contig of 21785 bp in length
110110 111119: gap of unknown length
111120 151295: contig of 40176 bp in length

FEATURES
source
1..231200
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-437E23"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 67796 a 43085 c 44997 g 73321 t 2001 others
ORIGIN

Query Match 6.7%; Score 50; DB 2; Length 231200;
Best Local Similarity 100.0%; Pred. No. 2.4e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 AATCTACTCCCTGGTGAAGATGCTGTGAACATTTGTTGAATGACAGAAG 176
|||||
Db 101782 AATCTACTCCCTGGTGAAGATGCTGTGAACATTTGTTGAATGACAGAAG 101733
|||||

RESULT 13
HSU49973
LOCUS 2418 bp DNA linear PRI 28-JUN-1997
DEFINITION Human Tigger1 transposable element, complete consensus sequence.
ACCESSION U49973
VERSION U49973.1 GI:2226003
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Smit,A.F. and Riggs,A.D.
TITLE Triggers and DNA transposon fossils in the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (4), 1443-1448 (1996)
MEDLINE 96202298
PUBMED 8643651
REFERENCE 2 (bases 1 to 2418)
AUTHORS Robertson,H.M.
TITLE Members of the pogo superfamily of DNA-mediated transposons in the human genome
JOURNAL Mol. Gen. Genet. 252 (6), 761-766 (1996)
MEDLINE 97074895
PUBMED 8917322
REFERENCE 3 (bases 1 to 2418)
AUTHORS Robertson,H.M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-1996) Hugh M. Robertson, Entomology, University of Illinois at Urbana-Champaign, 505 S. Goodwin, Urbana, IL 61801, USA

COMMENT On Jun 28, 1997 this sequence version replaced gi:1224064.
FEATURES
source
1..2418
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/transposon="Tigger1"
/note="consensus sequence based on 50 full-length genomic sequences"
repeat_region 1..13
/rpt_type="inverted"
425..1789
CDS
/note="ORF1; MER37; putative transposase similar to pogo element"
/citation=[1]
/codon_start=1
/protein_id="AAB61714.1"
/db_xref="GI:2226004"
/translation="MASKSSSRKSKRTSLTNOKLEMLKSEEGSKAEIGRKLGLLR
QTWSOVNAKRFLEIKSATPVNTRMIRKNSLIADMEKVLVVMIEDOTSHTNPLSQ
SLIQSKALTLENSMKAERGEAAEKELEASRGWFRFKERSRLNLIKVGGAASADGE
AAASPEDLAKIIDGGYKQIIFNVDETATFVKKMPSRTFIAREEKSMPPGFASKDR
LTLLEGANAAGDFKLKPLMIYHSENPALKNYAKSTLPVLYKNNKAMTAAHLPTAFW

TEYFKPTVTCSEKKISKIFILLIDNAPGHPRALMEMYKEINVVPANTTSILQPM
 DGGVISTKSYLNTFRKAIAAADSSDGSQKLTFWKGTIILDAIKNRDSWE
 EVKISTLGVWKKLIPTLMDDFEGFKTSVEEVTADEVIEARELEVEDEVTELLQS
 HDKT"

CDS

1811..2206
 /note="ORF2: function unknown"
 /codon_start=1
 /protein_id="BAB61715.1"
 /db_xref="GI:2226005"
 /translation="MDQRKFWLEWSTPGDANVIVMTTKDLEYINLVNKAAGF
 ERIDNFRSTVGRKMSNSIACYEIFRKRKQSMRQTSLLSYFKKLPPQPPQSATT
 TLISQPPSTKODPPPAKRLRLAEGSDDR"
 polya_signal 2218..2223
 repeat_region 2405..2418
 /rpt_type=inverted

BASE COUNT 746 a 477 c 529 g 666 t
 ORIGIN

Query Match 6.5%; Score 49; DB 9; Length 2418;
 Best Local Similarity 100.0%; Pred. No. 7.6e-17;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGAATCTACTCTCGTGAAGATGCTGTGACATTGTGAATGACAA 171
 |||||||

Db 1841 ATGGAATCTACTCTCGTGAAGATGCTGTGACATTGTGAATGACAA 1889
 |||||||

RESULT 14

AC022425 AC022425 92872 bp DNA linear PRI 02-NOV-2001
 LOCUS Homo sapiens chromosome 1 clone CTD-2302K4, complete sequence.
 DEFINITION

AC022425

AC022425

VERSION

AC022425.6 GI:14329094

KEYWORDS

HTG.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 92872)

Direct Submission

Unpublished

2 (bases 1 to 92872)

DOE Joint Genome Institute.

DOE Joint Genome Institute.

Direct Submission

Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 92872)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (28-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 92872)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (02-NOV-2001)

On Jun 8, 2001 this sequence version replaced gi:13470148.

Draft sequence produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.8% of Sequence;

Estimated Total Number of Errors is 0.2.

Location/Qualifiers

1..92872

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

FEATURES

source

BASE COUNT 26843 a 19114 c 18957 g 27958 t
 ORIGIN

Query Match 6.5%; Score 49; DB 9; Length 92872;
 Best Local Similarity 100.0%; Pred. No. 8.8e-17;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGAATCTACTCTCGTGAAGATGCTGTGACATTGTGAATGACAA 171
 |||||||

Db 13060 ATGGAATCTACTCTCGTGAAGATGCTGTGACATTGTGAATGACAA 13108
 |||||||

RESULT 15

AL589863

LOCUS

DEFINITION

Human DNA sequence from clone RP11-310J24 on chromosome 10,

complete sequence.

ACCESSION

AL589863

VERSION

AL589863.11 GI:17644185

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 168231)

Tracey, A.

Direct Submission

Submitted (10-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk

On Dec 12, 2001 this sequence version replaced gi:16973910.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; SW.,

SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/chr10

RP11-310J24 is from the library RP11-11.2 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-310J24 The true

right end of clone RP11-241117 is at 112163 in this sequence.

FEATURES

source

Location/Qualifiers

1..168231

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-310J24"

/clone.lib="RP11-11.2"

168148..168231

misc_feature

Location/Qualifiers

/note="Single clone region. Assembly confirmed by

restriction digest data."

BASE COUNT 46096 a 32857 c 35785 g 53493 t

ORIGIN

Query Match 6.5% Score 49; DB 9; Length 168231;
 Best Local Similarity 100.0%; Pred. No. 9e-17;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGAATCTACTCCTGGTGAAGATGCTGTGAACATTGTTGAAATGACAA 171
 |||||
 Db 163930 ATGGAATCTACTCCTGGTGAAGATGCTGTGAACATTGTTGAAATGACAA 163978

Search completed: April 12, 2003, 07:48:18
 Job time : 2405.57 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:38:14 : Search time 123.291 Seconds
(without alignments)
13717.504 Million cell updates/sec

Title: US-09-914-152-3_COPY_2000_2750

Perfect score: 751
Sequence: 1 cctgccttggtccccaag.....tgatggagctggaagtc 751

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	751	100.0	10562	21	AAA93876	Human beta3Gal-T5
2	46	6.1	1405	24	ABA05529	Molecular chaperon
3	42	5.6	32192	22	AAL04361	Human reproductive
4	38	5.1	525	23	AAS91985	DNA encoding novel
5	37	4.9	460	24	ABL65786	Lung cancer relate
6	36	4.8	3017	22	AAL14566	Human CDNA sequenc
7	35	4.7	199	21	AAC04612	Human secreted pro
8	35	4.7	361	22	AAC0367	Human polynucleoti
9	35	4.7	502	22	AAI86933	Human polynucleoti

10	35	4.7	536	22	AAH10509	Human cDNA clone (
11	35	4.7	1512	24	ABK34797	Human cDNA for nov
12	35	4.7	1701	22	AAH17571	Human cDNA sequenc
13	34	4.5	461	22	ABA58622	Human foetal liver
14	34	4.5	461	22	ABA27623	Probe #6089 for ge
15	34	4.5	461	22	AAK06749	Human brain expres
16	34	4.5	461	22	AAK32459	Human bone marrow
17	34	4.5	461	22	AAI16081	Probe #6014 for ge
18	34	4.5	461	22	AAI38294	Probe #6980 used t
19	34	4.5	461	24	ABS07243	Human genome-deriv
20	34	4.5	17700	22	AAK68945	Human immune/haema
21	34	4.5	21423	22	AAI36470	Human musculoskele
22	33	4.4	229	20	AAK51887	Human secreted pro
23	33	4.4	1677	23	AAK570746	DNA encoding novel
24	32	4.3	205	20	AAH85478	Human single nucle
25	32	4.3	205	20	AAH85479	Human single nucle
26	31	4.1	343	22	AAI91066	Human polynucleoti
27	31	4.1	21224	22	ABA20544	Human nervous syst
28	31	4.1	27976	22	AAK70187	Human immune/haema
29	31	4.1	27976	22	AAK79311	Human immune/haema
30	31	4.1	92638	24	ABQ88096	Human osteoblast d
31	30	4.0	420	22	ABA72957	Human foetal liver
32	30	4.0	420	22	ABA38514	Probe #16980 for g
33	30	4.0	420	22	AAK21388	Human brain expres
34	30	4.0	420	22	AAK47549	Human bone marrow
35	30	4.0	420	22	AAI25974	Probe #15907 for g
36	30	4.0	420	22	AAI53381	Probe #22067 used
37	30	4.0	420	24	ABS21646	Human genome-deriv
38	30	4.0	585	22	ABA60418	Human foetal liver
39	30	4.0	585	22	ABA28637	Probe #7103 for ge
40	30	4.0	585	22	AAK08696	Human brain expres
41	30	4.0	585	22	AAK34583	Human bone marrow
42	30	4.0	585	22	AAI16779	Probe #6712 for ge
43	30	4.0	585	22	AAI40302	Probe #8988 used t
44	30	4.0	585	24	ABS09214	Human genome-deriv
45	30	4.0	122888	24	ABR833569	Human cDNA differe

ALIGNMENTS

RESULT 1
AAA93876
ID AAA93876 standard; DNA; 10562 BP.
AC AAA93876;
DT 15-JAN-2001 (first entry)
DE Human beta3Gal-T5 encoding DNA.
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; beta3Gal-T5; ds.
OS Homo sapiens.
PN WO2000050608-Al.
PD 31-AUG-2000.
PF 24-FEB-2000; 2000WO-JP01070.
PR 25-FEB-1999; 99JP-0047571.
PA (KYOW) KYOWA HAKKO KOGYO KK.
PI Narimatsu H, Ishiki S, Todayachi A, Sasaki K;
DR WPI; 2000-549409/50.
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer

XX PS Claim 31; Page 103-111; 123pp; Japanese.

CC This invention relates to a polypeptide (I) with beta-1,3 galactose transferase activity, or variants of (I) comprising amino acid additions, deletions and/or substitutions. Included in the invention is DNA encoding all or part of (I); expression vectors containing the DNA, host cells transformed by the vectors; a method for the preparation of the polypeptide by culture of the transformants or by expression of the gene of a transgenic mammal, and antibodies recognising (I). The Beta-1,3 galactose transferase protein transfers galactose by beta-1,3 bonding to N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as GlcNAc-beta1-3galactal-4Glc) to give Galbeta1-3GlcNAc. The protein and DNA encoding it are useful for the treatment and diagnosis of cancer of the digestive system. The present sequence represents a Beta3gal-T5 encoding DNA sequence.

XX SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 100.0%; Score 751; DB 21; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGCTTGGGCTCCCAAGTGTAGATTACAGGGATGAGCCACCATGGTCACTTCAT 60
DB 2000 CCTGCTTGGGCTCCCAAGTGTAGATTACAGGGATGAGCCACCATGGTCACTTCAT 2059

QY 61 GATAAACTTCAGTGGATGAGAGCTGCCCTTATGATGAACAAAGAGTGTTCTTG 120
DB 2060 GATAAACTTCAGTGGATGAGAGCTGCCCTTATGATGAACAAAGAGTGTTCTTG 2119

QY 121 AAATGGAATCTACTCTGCTGGTGAAGATCTGTGAACATTTGTAATGACAAAGAAATT 180
DB 2 20 AAATGGAATCTACTCTGCTGGTGAAGATCTGTGAACATTTGTAATGACAAAGAAATT 2179

QY 181 TACAGTGTACATAGATAGTACTTGAAGCAGTAGCAGGATTCGAGAGATCGATTCCA 240
DB 2180 TACAGTGTACATAGATAGTACTTGAAGCAGTAGCAGGATTCGAGAGATCGATTCCA 2239

QY 241 ATTCAAATAGTCTTCTCTGGGTAATGCTATCAATGGGTCGCGATGACAGAG 300
DB 2240 ATTCAAATAGTCTTCTCTGGGTAATGCTATCAATGGGTCGCGATGACAGAG 2299

QY 301 AAATCTATCATGAAAGGAGTCAATGATGTCGCAAACTTCATTGTTGCTGATTTTA 360
DB 2300 AAATCTATCATGAAAGGAGTCAATGATGTCGCAAACTTCATTGTTGCTGATTTTA 2359

QY 361 AGAAATGTGAGGACCCCAACCTTCAACAAACCATGACCTGATCAGTCAGGACCAT 420
DB 2360 AGAAATGTGAGGACCCCAACCTTCAACAAACCATGACCTGATCAGTCAGGACCAT 2419

QY 421 CCACATTGAGGCGAGAACTCCAGCAGTAAAGATTTATGATCTCTTAAAGGATCAGATG 480
DB 2420 CCACATTGAGGCGAGAACTCCAGCAGTAAAGATTTATGATCTCTTAAAGGATCAGATG 2479

QY 481 ACATTAGCATTTTTTAAAGCAATTAAGTATTTTACGTAGATATGATGTTATTTTT 540
DB 2480 ACATTAGCATTTTTTAAAGCAATTAAGTATTTTACGTAGATATGATGTTATTTTT 2539

QY 541 AGGCATATGCTATTATGCTATTTATAGCTCCAGTATTTGTAACATTAACATTTAAATG 600
DB 2540 AGGCATATGCTATTATGCTATTTATAGCTCCAGTATTTGTAACATTAACATTTAAATG 2599

QY 601 CACTGGGAGATTAACATTTTCTCTTTTATGATATTTGCTTTATGCGATGCTCTGTA 660
DB 2600 CACTGGGAGATTAACATTTTCTCTTTTATGATATTTGCTTTATGCGATGCTCTGTA 2659

QY 661 TGGAACTACATTTCTCTTGGGACAGCTGTATACAGAAAGAAATTTATCATGAGGAAA 720
DB 2660 TGGAACTACATTTCTCTTGGGACAGCTGTATACAGAAAGAAATTTATCATGAGGAAA 2719

QY 721 TGCTCATCAATGATGGAGGCTGGAAAGTCC 751
DB 721 TGCTCATCAATGATGGAGGCTGGAAAGTCC 751

DB 2720 TGCTCATCAATGATGGAGGCTGGAAAGTCC 2750

RESULT 2

ABAO5529
ID ABAO5529 standard; cDNA; 1405 BP.
XX
AC ABAO5529;
XX
DT 25-FEB-2002 (first entry)
DE Molecular chaperone CPN60 protein 10 cDNA.
XX
XX Molecular chaperone; CPN60; protein 10; cytosolic; haemostatic;
KW virucide; immunomodulatory; antiinflammatory; gene therapy; cancer;
KW haemopathy; human immunodeficiency virus; HIV; infection;
KW immunological disease; inflammatory disorder; ss.

Unidentified.

XX Key Location/Qualifiers
XX 316...585
FT CDS
FT //tag= a
FT /product= 'CPN60 protein 10'

WO200190173-A1.

XX 29-NOV-2001.

XX 21-MAY-2001; 2001WO-CN00827.

XX 24-MAY-2000; 2000CN-0115833.

XX (SHAN-) SHANGHAI BROWNDOW GENE DEV INC.

XX Mao Y, Xie Y;

XX WPI: 2002-083089/11.

XX P-PSDB; AAMS1649.

XX Chaperone CPN60 protein 10 and encoding polynucleotide, used in diagnosis and treatment of malignant tumors, hemopathy, human immunodeficiency virus infection, immunological diseases and inflammation.

XX Claim 6; Page 29-30; 36pp; Chinese.

XX The invention relates to an isolated polypeptide of Chaperone CPN60 protein 10 comprising an 89 residue amino acid sequence, fully defined in the specification, or its fragment, analogue or derivative. The polypeptide is useful in the diagnosis and treatment of malignant tumors, haemopathy, human immunodeficiency virus (HIV) infection, immunological diseases and various inflammatory disorders. The present sequence encodes the polypeptide of the invention.

XX SQ Sequence 1405 BP; 460 A; 239 C; 312 G; 394 T; 0 other;

Query Match 6.1%; Score 46; DB 24; Length 1405;
Best Local Similarity 100.0%; Pred. No. 7.9e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 CAAGAAGGTTGTTCTTGAATGGAATCTACTCTCTGGTGAAGATG 147

DB 983 CAAGAAGGTTGTTCTTGAATGGAATCTACTCTCTGGTGAAGATG 1028

RESULT 3

AAL04361
ID AAL04361 standard; DNA; 32192 BP.

XX AAL04361;

XX 21-NOV-2001 (first entry)

XX	Human reproductive system related antigen DNA SEQ ID NO: 7049.	PR	14-SEP-2000;	2000US-0233065.
DE		PR	21-SEP-2000;	2000US-0234223.
XX		PR	21-SEP-2000;	2000US-0234273.
KW	Human; reproductive system related antigen; reproductive system disorder;	PR	25-SEP-2000;	2000US-0234997.
XX	cancer; gene therapy; ds.	PR	25-SEP-2000;	2000US-0234998.
XX		PR	26-SEP-2000;	2000US-0235484.
OS	Homo sapiens.	PR	27-SEP-2000;	2000US-0235834.
XX		PR	27-SEP-2000;	2000US-0235836.
PN	WO200155320-A2.	PR	29-SEP-2000;	2000US-0236327.
XX		PR	29-SEP-2000;	2000US-0236367.
PD	02-AUG-2001.	PR	29-SEP-2000;	2000US-0236368.
XX		PR	29-SEP-2000;	2000US-0236369.
XX		PR	29-SEP-2000;	2000US-0236370.
PF		PR	02-OCT-2000;	2000US-0236802.
XX		PR	02-OCT-2000;	2000US-0237037.
XX		PR	02-OCT-2000;	2000US-0237038.
PR		PR	02-OCT-2000;	2000US-0237039.
PR		PR	02-OCT-2000;	2000US-0237039.
PR		PR	02-OCT-2000;	2000US-0237040.
PR		PR	13-OCT-2000;	2000US-0239935.
PR		PR	13-OCT-2000;	2000US-0239937.
PR		PR	20-OCT-2000;	2000US-0240960.
PR		PR	20-OCT-2000;	2000US-0241221.
PR		PR	20-OCT-2000;	2000US-0241785.
PR		PR	20-OCT-2000;	2000US-0241786.
PR		PR	20-OCT-2000;	2000US-0241787.
PR		PR	20-OCT-2000;	2000US-0241808.
PR		PR	20-OCT-2000;	2000US-0241809.
PR		PR	20-OCT-2000;	2000US-0241826.
PR		PR	01-NOV-2000;	2000US-0244617.
PR		PR	08-NOV-2000;	2000US-0246474.
PR		PR	08-NOV-2000;	2000US-0246475.
PR		PR	08-NOV-2000;	2000US-0246476.
PR		PR	08-NOV-2000;	2000US-0246477.
PR		PR	08-NOV-2000;	2000US-0246478.
PR		PR	08-NOV-2000;	2000US-0246523.
PR		PR	08-NOV-2000;	2000US-0246524.
PR		PR	08-NOV-2000;	2000US-0246525.
PR		PR	08-NOV-2000;	2000US-0246526.
PR		PR	08-NOV-2000;	2000US-0246527.
PR		PR	08-NOV-2000;	2000US-0246528.
PR		PR	08-NOV-2000;	2000US-0246532.
PR		PR	08-NOV-2000;	2000US-0246609.
PR		PR	08-NOV-2000;	2000US-0246610.
PR		PR	08-NOV-2000;	2000US-0246611.
PR		PR	08-NOV-2000;	2000US-0246613.
PR		PR	17-NOV-2000;	2000US-0249207.
PR		PR	17-NOV-2000;	2000US-0249208.
PR		PR	17-NOV-2000;	2000US-0249209.
PR		PR	17-NOV-2000;	2000US-0249210.
PR		PR	17-NOV-2000;	2000US-0249211.
PR		PR	17-NOV-2000;	2000US-0249212.
PR		PR	17-NOV-2000;	2000US-0249213.
PR		PR	17-NOV-2000;	2000US-0249214.
PR		PR	17-NOV-2000;	2000US-0249215.
PR		PR	17-NOV-2000;	2000US-0249216.
PR		PR	17-NOV-2000;	2000US-0249217.
PR		PR	17-NOV-2000;	2000US-0249218.
PR		PR	17-NOV-2000;	2000US-0249244.
PR		PR	17-NOV-2000;	2000US-0249245.
PR		PR	17-NOV-2000;	2000US-0249264.
PR		PR	17-NOV-2000;	2000US-0249265.
PR		PR	17-NOV-2000;	2000US-0249297.
PR		PR	17-NOV-2000;	2000US-0249299.
PR		PR	01-DEC-2000;	2000US-0249300.
PR		PR	01-DEC-2000;	2000US-0250160.
PR		PR	05-DEC-2000;	2000US-0250391.
PR		PR	05-DEC-2000;	2000US-0251030.
PR		PR	05-DEC-2000;	2000US-0251988.
PR		PR	06-DEC-2000;	2000US-0256719.
PR		PR	08-DEC-2000;	2000US-0251479.
PR		PR	08-DEC-2000;	2000US-0251856.
PR		PR	08-DEC-2000;	2000US-0251868.
PR		PR	08-DEC-2000;	2000US-0251869.

```
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition
XX
XX Disclosure; SEQ ID NO 7049; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX
XX Sequence 32192 BP; 9426 A; 6338 C; 6511 G; 9917 T; 0 other;
SQ
Query Match 5.6%; Score 42; DB 22; Length 32192;
Best Local Similarity 100.0%; Pred. No. 7.7e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 CTCCTGGGAGAGCTGTGAACATTGTTGAAATGACAAGAA 174
DB 7637 CTCCTGGTGAAGATGCTGTGNACATTGTTGAATGACAAGAA 7678
RESULT 4
AAS91985
ID AAS91985 standard; cDNA; 525 BP.
XX
AC AAS91985;
XX
XX 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27789.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSDB; ABG27796.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders of other traits and to assess
XX biodiversity
XX
XX Claim 1; SEQ ID NO 27789; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC
```

```
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 525 BP; 204 A; 81 C; 136 G; 104 T; 0 other;
SQ
Query Match 5.1%; Score 38; DB 23; Length 525;
Best Local Similarity 100.0%; Pred. No. 8.1e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 123 ATGGAATCTACTCCTGGTGGAAGATGCTGTGAACATTGT 150
DB 133 ATGGAATCTACTCCTGGTGGAAGATGCTGTGAACATTGT 170
RESULT 5
ABL65786/c
ID ABL65786 standard; DNA; 460 BP.
XX
AC ABL65786;
XX
XX 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4123.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 22-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 23-SEP-2000; 2000US-234923P.
XX 23-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
```

PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 03-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 XX
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI; 2002-188264/24.
 DR
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set
 XX
 XX Claim 1; SEQ ID 4123; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 847 sequences (given in ABL61664
 CC to AB70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX
 SQ Sequence 460 BP; 114 A; 92 C; 94 G; 160 T; 0 other;
 Query Match 4.9%; Score 37; DB 24; Length 460;
 Best Local Similarity 100.0%; Pred. No. 2.6e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 135 CCTGGTGAAGATGCTGTGAACATGTTGAATGACAA 171
 Db 389 CCTGGTGAAGATGCTGTGAACATGTTGAATGACAA 353
 RESULT 6
 AAH14566
 ID AAH14566 standard; CDNA; 3017 BP.
 XX
 AC AAH14566;
 XX
 DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:12147.
 DE
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 PR
 XX 11-JAN-2000; 2000JP-0118776.
 PR
 XX 02-MAY-2000; 2000JP-0183767.
 PR
 XX 09-JUN-2000; 2000JP-0241899.
 PR
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 8; SEQ ID 12147; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 3017 BP; 895 A; 529 C; 711 G; 782 T; 0 other;
 Query Match 4.8%; Score 36; DB 22; Length 3017;
 Best Local Similarity 100.0%; Pred. No. 8e-08;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 114 TTCTTGAATGGAATCTACTCTCTGGTGAAGATGCT 149
 Db 1949 TTCTTGAATGGAATCTACTCTCTGGTGAAGATGCT 1984
 RESULT 7
 AAC04612/C
 ID AAC04612 standard; CDNA; 199 BP.
 XX
 AC AAC04612;
 XX
 DT

PA (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
DR P-PSDB; AAC07002.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX
XX Claim 1: SEQ ID NO 6993; 1399pp + Sequence Listing; English.
PS
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA00010-AA013910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 502 BP; 104 A; 114 C; 101 G; 177 T; 6 other;

Query Match 4.7%; Score 35; DB 22; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCGTTGGGCTCCCAAGTGTAGGATTACAGG 35
DB 171 CTGCGTTGGGCTCCCAAGTGTAGGATTACAGG 205
|||||

RESULT 10
AAH10509
ID AAH10509 standard; cDNA; 536 BP.
XX
XX AAH10509;
AC
XX
XX 26-JUN-2001 (first entry)
DE
XX Human cDNA clone (3'-primer) SEQ ID NO:7344.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS
XX EPI074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
PR
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
XX Claim 3: SEQ ID 7344; 2537pp + CD ROM; English.
PS
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to a
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 536 BP; 106 A; 145 C; 126 G; 153 T; 6 other;

Query Match 4.7%; Score 35; DB 22; Length 536;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCCAAAGTGTAGGATTACAGGATGAGCCACC 46
DB 235 CTCCCAAAGTGTAGGATTACAGGATGAGCCACC 269
|||||

RESULT 11
ABK34797/C
ID ABK34797 standard; cDNA; 1512 BP.
XX
XX ABK34797;
AC
XX
XX 08-MAY-2002 (first entry)
DE
XX Human cDNA for novel secreted protein, SEQ ID 566.
XX
XX Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
OS
XX Homo sapiens.
XX
XX WO200177290-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 29-MAR-2001; 2001WO-US10295.
PF
XX
XX 06-APR-2000; 2000US-194941P.
PR
XX
XX (GEMY) GENETICS INST INC.
PA
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;

XX WPI; 2002-179323/23.
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
XX human tissue sources which encode secreted proteins, useful for
XX treating immune deficiencies and disorders such as autoimmune disorders
XX
XX Claim 1; Page 303-304; 339pp; English.
XX
XX The invention relates to 625 polynucleotides which have been derived from
XX a variety of human tissue sources and which encode novel secreted
XX proteins, their complements and sequences that hybridise to them.
XX Also included are a vector comprising the polynucleotide, a host cell
XX transformed with the vector, the proteins encoded by the
XX polynucleotides, antibodies that bind to the proteins and identification
XX of modulators of the proteins or the expression of the polynucleotide.
XX The polynucleotides can be used as probes for the identification
XX and isolation of full length cDNA and genomic DNA. The polynucleotides
XX and proteins can also be used as nutritional supplements. The protein
XX is useful in the treatment of various immune deficiencies and disorders
XX such as viral infections, bacterial infections, fungal infections,
XX autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
XX autoimmune thyroiditis and diabetes) and allergic reactions and
XX conditions (e.g. asthma). They are also useful for treating
XX neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
XX disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
XX inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
XX useful for tissue regeneration, for wound healing and in the treatment
XX of burns, incisions and ulcers. The proteins are also useful for
XX regulating haematopoiesis, for treating myeloid or lymphoid cell
XX deficiencies. The present sequence is one of the 625 cDNA sequences
XX encoding a secreted protein.
XX
XX Sequence 1512 BP; 440 A; 340 C; 348 G; 384 T; 0 other;
SQ
Query Match 4.7%; Score 35; DB 24; Length 1512;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGCTGGGCTCCCAAGTCTAGGATTACAGG 35
|||||
Db 1282 CCGCTGGGCTCCCAAGTCTAGGATTACAGG 1248
|||||
RESULT 12
AAH1751/c
ID AAH17571 standard; cDNA; 1701 BP.
XX
XX AAH17571;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17057.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-030253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sudiama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 8; SEQ ID 17057; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 1701 BP; 513 A; 404 C; 417 G; 367 T; 0 other;
Query Match 4.7%; Score 35; DB 22; Length 1701;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 12 CTCCTCAAAGTCTAGGATTACAGGGATGAGCCACC 46
|||||
Db 1467 CTCCTCAAAGTCTAGGATTACAGGGATGAGCCACC 1433
|||||
RESULT 13
ABA58622/c
ID ABA58622 standard; DNA; 461 BP.
XX
XX ABA58622;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #6927.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX

XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
XX Claim 1; SEQ ID NO 6927; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 461 BP; 126 A; 114 C; 60 G; 161 T; 0 other;
Query Match 4.5%; Score 34; DB 22; Length 461;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 123 ATGGAATCTACTCTCGTGAAGATGCTGTGAACA 156
DB 213 ATGGAATCTACTCTCGTGAAGATGCTGTGAACA 180
RESULT 14
ABA27623/c
ID ABA27623 standard; DNA; 461 BP.
XX AC ABA27623;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #6089 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 1; SEQ ID NO 6089; 530pp; English.

XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 461 BP; 126 A; 114 C; 60 G; 161 T; 0 other;
Query Match 4.5%; Score 34; DB 22; Length 461;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 123 ATGGAATCTACTCTCGTGAAGATGCTGTGAACA 156
DB 213 ATGGAATCTACTCTCGTGAAGATGCTGTGAACA 180
RESULT 15
AAK06749/c
ID AAK06749 standard; DNA; 461 BP.
XX AC AAK06749;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 6740.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO: 6740; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.

XX

SQ Sequence 461 BP; 126 A; 114 C; 60 G; 161 T; 0 other;

Query Match

4.5%; Score 34; DB 22; Length 461;

Best Local Similarity 100.0%; Pred. No. 8.2e-07;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ATGGAATCTACTCTGCTGGAAGATGCTGTGAACA 156

|||||

Db 213 ATGGAATCTACTCTGCTGGAAGATGCTGTGAACA 180

Search completed: April 12, 2003, 03:53:42

Job time : 143.291 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 02:50:35 ; Search time 824.43 Seconds
(without alignments)
14752.993 Million cell updates/sec

Title: US-09-914-152-3_COPY_2000_2750

Perfect score: 751
Sequence: 1 cctgcctgggtcccaag.....tgatggaggctggaaagttcc 751

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_othr:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	239	31.8	318	10 AW859148	AW859148 MR1-CT035
2	138	18.4	209	10 BE000401	BE000401 MR0-BN007
3	137	18.2	196	10 AW999461	AW999461 MR0-BN007
4	46	6.1	448	9 AI001777	AI001777 of41f07 s
5	45	6.0	528	12 BF760554	BF760554 PM4-CT064
6	45	6.0	935	12 BG749551	BG749551 602707511

C	7	43	5.7	570	17	BI5851
	8	43	5.7	933	14	BQ959144
	9	42	5.6	453	17	AQ762388
	10	42	5.6	513	17	AQ581273
	11	42	5.6	557	17	AZ519146
	12	40	5.3	763	17	AQ747948
	13	39	5.2	652	17	AG153641
	14	38	5.1	406	17	AQ594909
	15	38	5.1	474	17	AQ383337
	16	38	5.1	496	17	AQ315231
C	17	38	5.1	525	17	AQ037355
	18	38	5.1	693	17	AQ037092
C	19	37	4.9	206	17	AQ016698
C	20	37	4.9	379	10	BE246065
C	21	37	4.9	415	9	AI144166
C	22	37	4.9	460	9	AA608733
C	23	37	4.9	465	14	N24942
	24	37	4.9	480	14	BM703722
	25	37	4.9	515	14	BM691815
C	26	37	4.9	558	13	BM668477
	27	37	4.9	574	17	AQ706497
C	28	37	4.9	914	9	AL521707
	29	37	4.9	994	12	BG113167
	30	36	4.8	391	9	AA232254
C	31	36	4.8	401	9	AA232751
C	32	36	4.8	404	14	R02724
C	33	36	4.8	418	14	H65341
C	34	36	4.8	423	9	AI240703
	35	36	4.8	462	17	AO685148
	36	36	4.8	477	14	N52421
C	37	36	4.8	535	10	AW889913
C	38	36	4.8	560	10	AW997051
C	39	36	4.8	574	12	BF433499
	40	36	4.8	656	13	EG982749
	41	36	4.8	805	17	AQ320297
	42	36	4.8	845	17	AQ747199
	43	35	4.7	179	9	AA689272
	44	35	4.7	335	9	AA372323
	45	35	4.7	368	13	BI012887

ALIGNMENTS

RESULT 1
AW859148/c 318 bp mRNA linear EST 19-MAY-2000
MR1-CT0352-220200-102-e01 CT0352 Homo sapiens cDNA, mRNA sequence.
AW859148
AW859148.1 GI:7954841
EST.
SOURCE
Homo sapiens
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 318)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR1-CT0352-220
 200-102-e01at3-2000-02-22at4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 318.

FEATURES

Location/Qualifiers
 1..318
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CT0352"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 91 a 77 c 45 g 105 t
 ORIGIN

Query Match 31.8%; Score 239; DB 10; Length 318;
 Best Local Similarity 99.7%; Pred. No. 8e-114;
 Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 54 ACTTCATGATAAACTTCAGTGGATGAGGAGCTGCCCTTTATGATGACAAAGAGGTGG 113
 Db 318 ACTTCATGATAAACTTCAGTGGATGAGGAGCTGCCCTTTATGATGACAAAGAGGTGG 259
 QY 114 TTCTTTGAATGGAAATCTACTCTCGTTCAGATGCTGTGGAACATTTGTTGAATGACAAGA 173
 Db 258 TTCTTTGAATGGAAATCTACTCTCGTTCAGATGCTGTGGAACATTTGTTGAATGACAAGA 199
 QY 174 AAGATTTACAGTGTTACATAGAGTTAGTTGATGAAGCAGTAGCAGGATTCGAGAGGATC 233
 Db 198 AAGATTTACAGTGTTACATAGAGTTAGTTGATGAAGCAGTAGCAGGATTCGAGAGGATC 139
 QY 234 GATTCCAATTTCAAAATTAAGTTCTTCGTGGTAAATGCTATCAAAATGCGTCGGATGC 293
 Db 138 GATTCCAATTTCAAAATTAAGTTCTTCGTGGTAAATGCTATCAAAATGCGTCGGATGC 79
 QY 294 TACAGAGAATCTATCTGAAGGAAGAGTCAATTGATGTGGCAAACTTC 343
 Db 78 TACAGAGAATCTATCTGAAGGAAGAGTCAATTGATGTGGCAAACTTC 29

RESULT 2
 BE000401
 LOCUS MR0-BN0070-090500-020-h08 BN0070 Homo sapiens cdna, mRNA sequence.
 DEFINITION
 ACCESSION BE000401
 VERSION BE000401.1 GI:8260634
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 209)

REFERENCE
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Ludwig Institute for Cancer Research
 Laboratory of Cancer Genetics
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR0-BN0070-090500-020-h08at3-2000-05-09at4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 23
 High quality sequence stop: 209.

FEATURES

Location/Qualifiers
 1..209
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BN0070"
 /dev_stage="Adult"
 /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 51 a 45 c 55 g 58 t
 ORIGIN

Query Match 18.4%; Score 138; DB 10; Length 209;
 Best Local Similarity 100.0%; Pred. No. 5.6e-61;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCGCTGGGCTCCCAAAGTCTAGGATTACAGGATGAGCCACCATGTCGACATTCAT 60
 Db 72 CTGCGCTGGGCTCCCAAAGTCTAGGATTACAGGATGAGCCACCATGTCGACATTCAT 131
 QY 61 GATAAACTTCAGTGGATGAGGAGCTCCTTTATGATGACAAAGAGGTGTTCTTG 120
 Db 132 GATAAACTTCAGTGGATGAGGAGCTCCTTTATGATGACAAAGAGGTGTTCTTG 191
 QY 121 AAATGGAATCTACTCCTG 138
 Db 192 AAATGGAATCTACTCCTG 209

RESULT 3
 AW999461/c
 LOCUS MR0-BN0070-240300-007-g09 BN0070 Homo sapiens cdna, mRNA sequence.
 DEFINITION
 ACCESSION AW999461
 VERSION AW999461.1 GI:8259695
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 196)

REFERENCE
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

```

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR0-BN0070-240
300-007-9094t3-2000-03-24&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 196.
Location/Qualifiers
1..196
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0070"
/dev_stage="Adult"
/note="Organ: breast; normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from OPESUES PCR (U.S. Letters Patent application
No.196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
54 a 53 c 41 g 48 t
BASE COUNT
ORIGIN
Query Match 18.2%; Score 137; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.9e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTGCTTGGGCTCCCAAGTCTAGGATTACAGGATGAGCCACCATGTCGACATCAT 60
Db 137 CTGCTTGGGCTCCCAAGTCTAGGATTACAGGATGAGCCACCATGTCGACATCAT 78

Qy 61 GATAAACTTCAGTGGATGAGGAGCTCCCTCTTATGATGAACAAGAAGTGTTTCCTTG 120
Db 77 GATAAACTTCAGTGGATGAGGAGCTCCCTCTTATGATGAACAAGAAGTGTTTCCTTG 18

Qy 121 AAATGGAACTCTACTCCT 137
Db 17 AAATGGAACTCTACTCCT 1

RESULT 4
AI001777/c AI001777 448 bp mRNA linear EST 27-AUG-1998
LOCUS o411f07.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1619365
DEFINITION 3' similar to TR:Q13338 Q13538 ORF2: FUNCTION UNKNOWN. ;, mRNA
sequence.
ACCESSION AI001777.1 GI:3202248
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 448)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2020663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4-CT0641-
21200-005-b094t3-2000-12-22&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 16
High quality sequence stop: 87.
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0641"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:

```

```

1..448
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1619365"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGCGCGCCAAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo.
122 a 90 c 78 g 158 t
BASE COUNT
ORIGIN
Query Match 6.1%; Score 46; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ATGGAATCTACTCTCGTGAAGATGCTGTGAACATTTGTTGAATGA 168
Db 330 ATGGAATCTACTCTCGTGAAGATGCTGTGAACATTTGTTGAATGA 285

RESULT 5
BF760554 528 bp mRNA linear EST 12-JAN-2001
LOCUS PM4-CT0641-221200-005-b09 CT0641 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF760554
ACCESSION BF760554
VERSION BF760554.1 GI:12108454
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 528)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2020663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4-CT0641-
21200-005-b094t3-2000-12-22&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 16
High quality sequence stop: 87.
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0641"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:

```

SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 176 a 92 c 130 g 130 t
ORIGIN

Query Match 6.0%; Score 45; DB 12; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.7e-12;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGAATCTACTCTCGTGAAGATGCTGTGAACATGTTGAATG 167

|||||

Db 324 ATGAATCTACTCTCGTGAAGATGCTGTGAACATGTTGAATG 368

RESULT 6

BG749551/C

LOCUS 602707511F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844139 5',

DEFINITION mRNA sequence.

ACCESSION BG749551

VERSION BG749551

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 935)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999);

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI691 row: e column: 04

High quality sequence stop: 150.

Location/Qualifiers

1..935

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4844139"

/clone_lib="NIH_MGC_43"

/tissue_type="normal pigmented retinal epithelium"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library. |"

BASE COUNT 352 a 247 c 203 g 133 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2.7e-12;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GGTTCCTTGAATGAATCTACTCTGCTGAAGATGCTGTGAACA 156

|||||

Db 72 GGTTCCTTGAATGAATCTACTCTGCTGAAGATGCTGTGAACA 28

RESULT 7

B15851/C

LOCUS 346E17_TP CIT978SKA1 Homo sapiens genomic clone A-346E17, DNA

DEFINITION sequence.

ACCESSION B15851

VERSION B15851.1

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 570)

AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.

TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT Other_GSSs: 346E17_TV

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..570

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="A-346E17"

/clone_lib="CIT978SKA1"

/sex="Female"

/cell_type="Fibroblast"

/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;

Caltech Human BAC Library A1"

BASE COUNT 184 a 100 c 99 g 187 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 3e-11;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 TGAAGATGCTGTGAACATTTGTTGAATGACAAAGAATTTA 182

|||||

Db 523 TGAAGATGCTGTGAACATTTGTTGAATGACAAAGAATTTA 481

RESULT 8

BQ959144

LOCUS 346E17_TP CIT978SKA1 Homo sapiens genomic clone A-346E17, DNA

DEFINITION sequence.

ACCESSION BQ959144

VERSION BQ959144.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 933)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L102653 Row: f Column: 14
 High quality sequence start: 32
 High quality sequence stop: 575.
 Location/Qualifiers

FEATURES

source

1. .933

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6462829"
 /clone_lib="NIH_MGC 101"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pONB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
 BASE COUNT 282 a 169 c 241 g 241 t
 ORIGIN

Query Match 5.7%; Score 43; DB 14; Length 933;

Best Local Similarity 100.0%; Pred. No. 3e-11;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGAATCTACTCTGCTGAGATGCTGTGACATTGTTGAAA 165

DB 226 ATGGAATCTACTCTGCTGAGATGCTGTGACATTGTTGAAA 268

RESULT 9

AQ762388

LOCUS

DEFINITION HS_3190_A2_F04_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3190 Col=8 Row=K, DNA sequence.

ACCESSION AQ762388

VERSION AQ762388.1 GI:5640504

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 453)

MAHAIRAS,G.G., Wallace,J.C., Smith,K., Swartzell,S., Adams,M.D., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Zhao,S., and Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380389

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 3190 Row: K Column: 8

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 453.

Location/Qualifiers

1. .453

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=3190 Col=8 Row=K"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

FEATURES

source

1. .453

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=3190 Col=8 Row=K"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 158 a 63 c 111 g 119 t 2 others

ORIGIN

Query Match 5.6%; Score 42; DB 17; Length 453;

Best Local Similarity 100.0%; Pred. No. 9.9e-11;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GGAATCTACTCTGCTGAGATGCTGTGACATTGTTGAAAAT 166

DB 280 GGAATCTACTCTGCTGAGATGCTGTGACATTGTTGAAAAT 321

RESULT 10

AQ581273

LOCUS

DEFINITION AQ581273 513 bp DNA linear GSS 07-JUN-1999
 RPCI-11-79C8.TJC RPCI-11 Homo sapiens genomic clone RPCI-11-79C8,
 DNA sequence.

ACCESSION AQ581273

VERSION AQ581273.1 GI:5008383

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 513)

ZHAO,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished (1997)

Contact: RPCI11-79C8.TJ RPCI11-79C8.TV

Other_GSSs: RPCI11-79C8.TJ RPCI11-79C8.TV

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.

Seq primer: SB6

Class: BAC ends.

Location/Qualifiers

1. .513

/organism="Homo sapiens"

/db_xref="GDB:753007"

/db_xref="taxon:9606"

/clone="RPCI-11-79C8"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3 6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 189 a 68 c 107 g 148 t 1 others

ORIGIN

Query Match 5.6%; Score 42; DB 17; Length 513;

Best Local Similarity 100.0%; Pred. No. 1e-10;

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 CTACTCTGCTGAGATGCTGTGACATTGTTGAAATGACAA 171

DB 180 CTACTCTGCTGAGATGCTGTGACATTGTTGAAATGACAA 221

RESULT 11

```

A2519146
LOCUS      A2519146      557 bp      DNA      linear      GSS 16-OCT-2000
DEFINITION RPCI-11-79C8.TJD RPCI-11 Homo sapiens genomic clone RPCI-11-79C8,
DNA sequence.
ACCESSION  A2519146
VERSION     A2519146.1 GI:10830360
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 557)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
            BAC end sequences of library RPCI-11
            RPCI-11-79C8.TVC
            Other GSSs: RPCI11-79C8.TJ RPCI-11-79C8.TJC RPCI11-79C8.TV
            RPCI-11-79C8.TVC
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humden/bac_end_search/bac_end_search.html.
            This BAC end was generated during the R&B process and may have
            higher chance of clone tracking errors.
            Seq primer: SP6
            Class: BAC ends.
FEATURES    source
            Location/Qualifiers
                1..557
                /organism="Homo sapiens"
                /db_xref="GDB:7530007"
                /db_xref="taxon:9606"
                /clone="RPCI-11-79C8"
                /clone_lib="RPCI-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
                RPCI11 Human Male BAC Library"
BASE COUNT  195 a 85 c 113 g 164 t
ORIGIN
Query Match 5.6%; Score 42; DB 17; Length 557;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 CTACTCCTGGTGAAGATGCTGTGAACATTGTTGAATGACAA 171
      |||||||
DB 98 CTACTCCTGGTGAAGATGCTGTGAACATTGTTGAATGACAA 139
      |||||||

RESULT 12
AQ747948
LOCUS      AQ747948      763 bp      DNA      linear      GSS 19-JUL-1999
DEFINITION HS_5536_A2_H03.SP6 RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1112 Col-6 Row-O, DNA sequence.
ACCESSION  AQ747948
VERSION     AQ747948.1 GI:5535106
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 763)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Kaller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and

```

```

Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1112 row: 0 column: 6
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 763.
FEATURES    source
            Location/Qualifiers
                1..763
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate-1112 Col-6 Row-O"
                /clone_lib="RPCI-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
                Male blood DNA was isolated from one randomly chosen donor
                and partially digested with a combination of EcoRI and
                EcoRI Methylase. Size selected DNA was cloned into the
                pBACE3.6 vector at EcoRI sites"
BASE COUNT  257 a 120 c 156 g 227 t 3 others
ORIGIN
Query Match 5.3%; Score 40; DB 17; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 CTACTCCTGGTGAAGATGCTGTGAACATTGTTGAATGAC 169
      |||||||
DB 401 CTACTCCTGGTGAAGATGCTGTGAACATTGTTGAATGAC 440
      |||||||

RESULT 13
AG153641
LOCUS      AG153641      652 bp      DNA      linear      GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-017G17.T7, genomic survey
sequence.
ACCESSION  AG153641
VERSION     AG153641.1 GI:16683319
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            BAC Library clone:RP43-017G17.T7.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
            BAC end sequences of Library RPCI-43
            Unpublished
            2 (bases 1 to 652)
            Direct Submission
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22, Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
            {E-mail:chmpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170}

```

COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS Sequencing: T7

LIBRARY Vector : pBACE3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1. 652
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-017617.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
222 a 114 c 128 g 187 t 1 others

BASE COUNT 222 a 114 c 128 g 187 t 1 others

ORIGIN

Query Match 5.2%; Score 39; DB 17; Length 652;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGATCTACTCCTGGTGAAGATGCTGTGACATTGTT 161
|||||

Db 225 ATGGAATCTACTCCTGGTGAAGATGCTGTGACATTGTT 263
|||||

RESULT 14
AQ594909
LOCUS
DEFINITION HS-5426-AL-G07-SP6E RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION AQ594909
VERSION AQ594909.1 GI:5026495
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 406)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequencing-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1002 row: M column: 13
Seq primer: SP6
Class: BAC ends

High quality sequence stop: 406.
Location/Qualifiers
1. 406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1002 Col=13 Row=M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"

FEATURES
source

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 141 a 46 c 53 g 154 t 2 others

ORIGIN

Query Match 5.1%; Score 38; DB 17; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 ATGGAATCTACTCCTGGTGAAGATGCTGTGACATTGT 160
|||||

Db 335 ATGGAATCTACTCCTGGTGAAGATGCTGTGACATTGT 372
|||||

RESULT 15
AQ383337
LOCUS
DEFINITION RPC111-136B21-TV RPCI-11 Homo sapiens genomic clone RPCI-11-136B21, DNA linear GSS 21-MAY-1999
ACCESSION AQ383337
VERSION AQ383337.1 GI:4354360
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 474)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPC111-136B21.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/db_xref="GDB:755184"
/db_xref="taxon:9606"
/clone="RPCI-11-136B21"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"

FEATURES
source

BASE COUNT 155 a 55 c 110 g 154 t

Query Match 5.1%; Score 38; DB 17; Length 474;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TCTGTGTGAAGATGCTGTGACATTGTTGAAATGACAA 171
|||||

Db 294 TCTGTGTGAAGATGCTGTGACATTGTTGAAATGACAA 331
|||||

Search completed: April 12, 2003, 08:59:07
Job time : 833.43 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 03:59:25 ; Search time 20.8447 Seconds
(without alignments)
11049.076 Million cell updates/sec

Title: US-09-914-152-3_COPY_2000_2750
Perfect score: 751
Sequence: 1 cctgcctgggctcccaag.....tgatggaggctggaagttcc 751

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 1533381 residues

Word size : 0
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
C 1	33	4.4	229	4 US-08-905-223-138
C 2	28	3.7	2418	4 US-09-285-379-1
C 3	25	3.3	548	4 US-08-991-789A-15
4	25	3.3	548	4 US-09-082-451-15
5	25	3.3	548	4 US-08-598-326-15
6	25	3.3	567	1 US-08-480-784-30
7	25	3.3	567	1 US-08-483-553-30
8	25	3.3	567	1 US-08-487-002-30
9	25	3.3	567	1 US-08-483-554B-30
10	25	3.3	567	1 US-08-488-011B-30
11	25	3.3	567	4 US-08-850-727-30
12	25	3.3	567	5 PCT-US95-10202-30
13	25	3.3	567	5 PCT-US95-10203-30
14	25	3.3	567	5 PCT-US95-10220-30
C 15	24	3.2	547	4 US-09-085-199B-27
C 16	24	3.2	615	4 US-09-385-982-528
C 17	24	3.2	630	4 US-09-342-681C-5
C 18	24	3.2	680	4 US-09-227-357-144
C 19	24	3.2	850	4 US-09-288-143-26
C 20	24	3.2	901	1 US-08-832-883-65
C 21	24	3.2	901	2 US-08-832-877-65
C 22	24	3.2	931	4 US-09-227-357-113
C 23	24	3.2	999	4 US-09-345-882-3
C 24	24	3.2	1001	4 US-09-641-638-78
C 25	24	3.2	1001	4 US-09-641-638-376
C 26	24	3.2	1232	4 US-09-345-293-1
C 27	24	3.2	1260	1 US-08-599-252-83

C 28	24	3.2	1260	1 US-08-436-074-56
C 29	24	3.2	1260	5 PCT-US96-06352-83
C 30	24	3.2	1260	5 PCT-US96-06583-83
C 31	24	3.2	1287	4 US-09-564-805-217
C 32	24	3.2	1586	1 US-08-461-244-1
C 33	24	3.2	1630	4 US-09-175-928-17
C 34	24	3.2	2337	3 US-09-156-253-1
C 35	24	3.2	2477	1 US-08-429-742-1
C 36	24	3.2	4233	4 US-09-056-105-27
C 37	24	3.2	4576	1 US-08-832-883-49
C 38	24	3.2	4576	2 US-08-832-877-49
C 39	24	3.2	5543	2 US-08-687-080-101
C 40	24	3.2	5761	1 US-07-749-001-2
C 41	24	3.2	5761	1 US-08-154-198-2
C 42	24	3.2	5761	1 US-08-463-335-2
C 43	24	3.2	5761	2 US-08-464-023A-2
C 44	24	3.2	11613	1 US-08-484-044-10
C 45	24	3.2	13158	2 US-08-687-080-105

ALIGNMENTS

RESULT 1
US-08-905-223-138/c
: Sequence 138, Application US/08905223
: Patent No. 6222029
: GENERAL INFORMATION:
: APPLICANT: Edwards, Jean-Baptiste D.
: APPLICANT: Duclert, Aymeric
: APPLICANT: Lacroix, Bruno
: TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
: NUMBER OF SEQUENCES: 503
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 501 West Broadway
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-3505
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Win95
: SOFTWARE: Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,223
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Israel, Ned A.
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 138:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 229 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: DOUBLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: CDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo Sapiens
: TISSUE TYPE: Brain
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 101..157
: IDENTIFICATION METHOD: Von Heijne matrix
: OTHER INFORMATION: score 4.1
: OTHER INFORMATION: seq FVPESTMTASSP/GE

US-08-905-223-138

Query Match 4.4%; Score 33; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.4e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTCCTGGTGAAGATGCTGTGAACATGTTGAAA 165
|||||
DB 161 CTCCTGGTGAAGATGCTGTGAACATGTTGAAA 129

RESULT 2

US-09-285-379-1/c

; Sequence 1, Application US/09285379

; Patent No. 6451765

; GENERAL INFORMATION:

; APPLICANT: Ervin Jr., Paul R.

; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE OF MAMMASTATIN AND

; FILE REFERENCE: 4273.1USD1

; CURRENT APPLICATION NUMBER: US/09/285,379

; CURRENT FILING DATE: 1999-04-02

; EARLIER APPLICATION NUMBER: 08/943,828

; EARLIER FILING DATE: 1997-10-03

; EARLIER APPLICATION NUMBER: 60/027,315

; EARLIER FILING DATE: 1996-10-03

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; TYPE: DNA

; LENGTH: 2418

; ORGANISM: Homo sapiens

US-09-285-379-1

Query Match 3.7%; Score 28; DB 4; Length 2418;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCAAAGTGTAGGATACAGGGATGAGC 42
|||||

DB 2150 CCAAAGTGTAGGATACAGGGATGAGC 2123
|||||

RESULT 3

US-08-991-789A-15

; Sequence 15, Application US/08991789A

; Patent No. 6225054

; GENERAL INFORMATION:

; APPLICANT: Frudakis, Tony N.

; Smith, John M.

; Reed, Steven G.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TREATMENT AND DIAGNOSIS OF BREAST CANCER

; NUMBER OF SEQUENCES: 292

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed Ip Law Group

; STREET: 701 Fifth Avenue, Suite 6300

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/991,789A

; FILING DATE: 11-Dec-1997

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Potter, Jane E. R.

; REGISTRATION NUMBER: 33,332

; REFERENCE/DOCKET NUMBER: 210121.419C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 548 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 15:

US-08-991-789A-15

Query Match 3.3%; Score 25; DB 4; Length 548;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGCTCCCAAAGTGTAGGATTACAG 34
|||||

DB 469 GGCTCCCAAAGTGTAGGATTACAG 493
|||||

RESULT 4

US-09-062-451-15

; Sequence 15, Application US/09062451

; Patent No. 6344550

; GENERAL INFORMATION:

; APPLICANT: Frudakis, Tony N.

; APPLICANT: Smith, John M.

; APPLICANT: Reed, Steven G.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER

; NUMBER OF SEQUENCES: 297

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/062,451

; FILING DATE: 04-APR-1997

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.419C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 548 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-062-451-15

Query Match 3.3%; Score 25; DB 4; Length 548;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGCTCCCAAAGTGTAGGATTACAG 34
|||||

DB 469 GGCTCCCAAAGTGTAGGATTACAG 493
|||||

RESULT 5

GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington

RESULT 7
US-08-483-553-30
Sequence 30, Application US/08483553
Patent No. 5705999
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
Susceptibility Gene

NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-553-30

Query Match 3.3%; Score 25; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCTCAAGTGTAGGATTACAGG 36
|||||
Db 94 CTCCTCAAGTGTAGGATTACAGG 118

RESULT 8
US-08-487-002-30
Sequence 30, Application US/08487002
Patent No. 5710001
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer Susceptibility Gene

NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-30

Query Match 3.3%; Score 25; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCTCAAGTGTAGGATTACAGG 36
|||||
Db 94 CTCCTCAAGTGTAGGATTACAGG 118

RESULT 9
US-08-483-554B-30
Sequence 30, Application US/08483554B
Patent No. 5747282
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,554B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-554B-30

Query Match 3.3%; Score 25; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCTCAAGTCTAGGATTACAGG 36
|||||
DB 94 CTCCTCAAGTCTAGGATTACAGG 118

RESULT 10
US-08-488-011B-30
Sequence 30, Application US/08488011B
Patent No. 5753441
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.

APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,011B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-011B-30

Query Match 3.3%; Score 25; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCTCAAGTCTAGGATTACAGG 36
|||||
DB 94 CTCCTCAAGTCTAGGATTACAGG 118

RESULT 11
US-08-850-727-30

? Sequence 30, Application US/08850727
? Patent No. 6162897
? GENERAL INFORMATION:
? APPLICANT: Skolnick, Mark H.
? APPLICANT: Goldgar, David E.
? APPLICANT: Miki, Yoshio
? APPLICANT: Swenson, Jeff
? APPLICANT: Kamb, Alexander
? APPLICANT: Harshman, Keith D.
? APPLICANT: Shattuck-Eidens, Donna M.
? APPLICANT: Tavtiglian, Sean V.
? APPLICANT: Wiseman, Roger W.
? APPLICANT: Futreal, P. Andrew
? TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
? TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
? TITLE OF INVENTION: Susceptibility Gene
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
? STREET: 1201 New York Avenue, N.W., Suite 1000
? CITY: Washington
? STATE: DC
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/850,727
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/483,554
? FILING DATE: 07-JUN-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/348,824
? FILING DATE: 29-NOV-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/308,104
? FILING DATE: 16-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/300,266
? FILING DATE: 02-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/289,221
? FILING DATE: 12-AUG-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Ihnen, Jeffrey L.
? REGISTRATION NUMBER: 28,957
? REFERENCE/DOCKET NUMBER: 24884-109347
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-962-4810
? TELEFAX: 202-962-8300
? INFORMATION FOR SEQ ID NO: 30:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 567 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
US-08-850-727-30

Query Match 3.3%; Score 25; DB 4; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 CTCCTCAAGTGTAGGATTACAGG 36
|||||

Db 94 CTCCTCAAGTGTAGGATTACAGG 118
RESULT 12
PCT-US95-10202-30
? Sequence 30, Application PC/TUS9510202
? GENERAL INFORMATION:
? APPLICANT: Shattuck-Eidens, Donna M.
? APPLICANT: Simard, Jacques
? APPLICANT: Eml, Mitsuru
? APPLICANT: Nakamura, Yusuke
? APPLICANT: Durocher, Francine
? TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
? TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer
? TITLE OF INVENTION: Susceptibility Gene
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
? STREET: 1201 New York Avenue, N.W., Suite 1000
? CITY: Washington
? STATE: DC
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US95/10202
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US
? FILING DATE: 07-JUN-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/409,305
? FILING DATE: 24-MAR-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/348,824
? FILING DATE: 29-NOV-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08-308,104
? FILING DATE: 16-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/300,266
? FILING DATE: 02-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/289,221
? FILING DATE: 12-AUG-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Ihnen, Jeffrey L.
? REGISTRATION NUMBER: 28,957
? REFERENCE/DOCKET NUMBER: 24884-109347
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-962-4810
? TELEFAX: 202-962-8300
? INFORMATION FOR SEQ ID NO: 30:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 567 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
PCT-US95-10202-30

Query Match 3.3%; Score 25; DB 5; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
PCT-US95-10202-30

QY 12 CTCCTCAAGTCTAGGATTACAGG 36
Db 94 CTCCTCAAGTCTAGGATTACAGG 118

RESULT 13

PCT-US95-10203-30

; Sequence 30, Application PC/TUS9510203

; GENERAL INFORMATION:

; APPLICANT: Skolnick, Mark H.

; APPLICANT: Goldgar, David E.

; APPLICANT: Miki, Yoshio

; APPLICANT: Swenson, Jeff

; APPLICANT: Kamb, Alexander

; APPLICANT: Harshman, Keith D.

; APPLICANT: Shattuck-Eidens, Donna M.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Wiseman, Roger W.

; APPLICANT: Futreal, P. Andrew

; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer

; TITLE OF INVENTION: Susceptibility Gene

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

; STREET: 1201 New York Avenue, N.W., Suite 1000

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10203

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/409,305

; FILING DATE: 24-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/348,824

; FILING DATE: 29-NOV-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08-308,104

; FILING DATE: 16-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/300,266

; FILING DATE: 02-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/289,221

; FILING DATE: 12-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957

; REFERENCE/DOCKET NUMBER: 24884-109347

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-962-4810

; TELEFAX: 202-962-8300

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 567 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

PCT-US95-10203-30

Query Match

Best Local Similarity 3.3%; Score 25; DB 5; Length 567;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCTCAAGTCTAGGATTACAGG 36

Db 94 CTCCTCAAGTCTAGGATTACAGG 118

RESULT 14

PCT-US95-10220-30

; Sequence 30, Application PC/TUS9510220

; GENERAL INFORMATION:

; APPLICANT: Skolnick, Mark H.

; APPLICANT: Goldgar, David E.

; APPLICANT: Miki, Yoshio

; APPLICANT: Swenson, Jeff

; APPLICANT: Kamb, Alexander

; APPLICANT: Harshman, Keith D.

; APPLICANT: Shattuck-Eidens, Donna M.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Wiseman, Roger W.

; APPLICANT: Futreal, P. Andrew

; TITLE OF INVENTION: Method for Diagnosing a

; TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

; STREET: 1201 New York Avenue, N.W., Suite 1000

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10220

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/409,305

; FILING DATE: 24-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/348,824

; FILING DATE: 29-NOV-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08-308,104

; FILING DATE: 16-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/300,266

; FILING DATE: 02-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/289,221

; FILING DATE: 12-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957

; REFERENCE/DOCKET NUMBER: 24884-109347

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-962-4810

; TELEFAX: 202-962-8300

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 567 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US95-10220-30

Query Match          3.3%; Score 25; DB 5; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCCAAAGTGCTAGGATTACAGG 36
    |||||
Db 94 CTCCCAAAGTGCTAGGATTACAGG 118

RESULT 15
US-09-085-1998-27/c
; Sequence 27, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Maibubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC-P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE:
; OTHER INFORMATION: exon 12 of HIP1
; US-09-085-1998-27

Query Match          3.2%; Score 24; DB 4; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.02;
```

```

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTCCCAAAGTGCTAGGATTACAGG 35
    |||||
Db 380 CTCCCAAAGTGCTAGGATTACAGG 357

Search completed: April 12, 2003, 09:02:57
Job time : 27.8447 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:52:34 ; Search time 1358.57 seconds
(without alignments)
16087.603 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750
Perfect score: 751
Sequence: 1 aaagcagactctagactca.....gggaggtctacacgttctgc 751

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba:
- 2: gb_htg:
- 3: gb_in:
- 4: gb_cm:
- 5: gb_cv:
- 6: gb_pat:
- 7: gb_ph:
- 8: gb_pl:
- 9: gb_pr:
- 10: gb_ro:
- 11: gb_sts:
- 12: gb_sy:
- 13: gb_un:
- 14: gb_vi:
- 15: em_ba:
- 16: em_fun:
- 17: em_hum:
- 18: em_in:
- 19: em_mu:
- 20: em_or:
- 21: em_ov:
- 22: em_pat:
- 23: em_ph:
- 24: em_pi:
- 25: em_ro:
- 26: em_sts:
- 27: em_un:
- 28: em_vi:
- 29: em_htg_hum:
- 30: em_htg_inv:
- 31: em_htg_mus:
- 32: em_htg_other:
- 33: em_htg_pln:
- 34: em_htg_rtd:
- 35: em_htg_nam:
- 36: em_htg_vrt:
- 37: em_sy:
- 38: em_hgo_hum:
- 39: em_hgo_mus:
- 40: em_hgo_other:
- 41: em_hgo_hum:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	751	100.0	10562	6	E38420	Novel polyp
2	751	100.0	170121	9	AF064860	AF064860 Homo sapi
3	751	100.0	340000	9	HS21C080	AL163280 Homo sapi
4	480	63.9	185982	2	AC073231	AC073231 Homo sapi
5	24	3.2	251899	2	AC122463	AC122463 Mus muscu
6	22	2.9	130815	9	AC007237	AC007237 Homo sapi
7	22	2.9	133526	2	AC105371	AC105371 Sus scrofa
8	22	2.9	146487	2	AC125568	AC125568 Rattus no
9	22	2.9	170928	9	AL356740	AL356740 Human DNA
10	22	2.9	211430	10	AL731778	AL731778 Mouse DNA
11	21	2.8	106578	9	AL137249	AL137249 Human DNA
12	21	2.8	110000	2	AC119727	AC119727 Homo sapi
13	21	2.8	135431	9	AC106847	AC106847 Homo sapi
14	21	2.8	136417	2	AC128939	AC128939 Rattus no
15	21	2.8	150507	9	AC107918	AC107918 Homo sapi
16	21	2.8	178902	2	AC061998	AC061998 Homo sapi
17	21	2.8	215734	2	AC073710	AC073710 Mus muscu
18	20	2.7	1347	6	AX004859	AX004859 Sequence
19	20	2.7	1911	9	S80071	S80071 nPro1-brain
20	20	2.7	2722	10	RATLPTHA	M88111 Rattus norv
21	20	2.7	2728	6	I30479	I30479 Sequence 7
22	20	2.7	2882	9	HS010277	AJ010277 Homo sapi
23	20	2.7	3738	9	AK096607	AK096607 Homo sapi
24	20	2.7	47160	2	AC101175	AC101175 Mus muscu
25	20	2.7	61658	2	AC101091	AC101091 Mus muscu
26	20	2.7	62736	2	AC102774	AC102774 Mus muscu
27	20	2.7	87857	9	AC005895	AC005895 Homo sapi
28	20	2.7	94763	2	AC126165	AC126165 Rattus no
29	20	2.7	106319	9	HS747L4	AL009051 Homo sapi
30	20	2.7	132639	9	AP000679	AP000679 Homo sapi
31	20	2.7	134137	9	HS3E5	AL022239 Human DNA
32	20	2.7	139164	2	AC120068	AC120068 Rattus no
33	20	2.7	140306	9	HS166816	AL023913 Human DNA
34	20	2.7	145357	2	AC121596	AC121596 Mus muscu
35	20	2.7	145814	2	AC122308	AC122308 Mus muscu
36	20	2.7	149194	9	AC007566	AC007566 Homo sapi
37	20	2.7	151447	3	AC010122	AC010122 Drosophila
38	20	2.7	152313	9	AL591594	AL591594 Human DNA
39	20	2.7	154865	9	AC099757	AC099757 Homo sapi
40	20	2.7	162247	2	AC090100	AC090100 Homo sapi
41	20	2.7	162820	2	AC019845	AC019845 Drosophila
42	20	2.7	163314	2	AL590557	AL590557 Homo sapi
43	20	2.7	169263	2	AC094047	AC094047 Rattus no
44	20	2.7	172862	2	AL645796	AL645796 Homo sapi
45	20	2.7	172862	2	AL645796	AL645796 Homo sapi

ALIGNMENTS

RESULT 1
E38420
LOCUS E38420 Novel polypeptide.
DEFINITION E38420 Novel polypeptide.
ACCESSION E38420
VERSION E38420.1 GI:13626994
KEYWORDS JP 2000245464-A/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 10562)
Narimatsu, H., Isshiki, S., Togayachi, A. and Sasaki, K.
TITLE Novel polypeptide
JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;

E38420 10562 bp DNA linear PAT 31-JAN-2002

COMMENT KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464-A/2
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR
PI HIGASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
C12P21/02,
PC C12P21/08, C12Q1/68, G01N33/53, C12N1/21, C12R1:185, C12N5/10,
C12R1:91,
PC (C12P21/02, C12R1:185), (C12P21/02, C12R1:91), C12N15/00, C12N5/00,
PC (C12N5/00, C12R1:91)
CC
FH Location/Qualifiers
FT promoter (1) (5000)
FT exon (5001) (5140)
FT exon (5001) (5273)
FT exon (5459) (5567)
FT exon (7427) (7586)
FT exon (8234) (10562)
FEATURES
source
1. 10562
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2610 a 2415 c 2574 g 2963 t
ORIGIN
Query Match 100.0%; Score 751; DB 6; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAGCAGACTCTAGCAGCTAGGACTGCGACACCCCTCCCGGGTTTCCTGTGCTGGAAT 60
DB 4000 AAAGCAGACTCTAGCAGCTAGGACTGCGACACCCCTCCCGGGTTTCCTGTGCTGGAAT 4059
QY 61 TCCCATCCCTGGTTCACCTGTTATCATCACACTCCCTTCAAGGACAGTGCAGATGC 120
DB 4060 TCCCATCCCTGGTTCACCTGTTATCATCACACTCCCTTCAAGGACAGTGCAGATGC 4119
QY 121 CAGTCTCCACGGGCTCAGATGTCACAGCTTCCTCCACCGAGGCGCACAGCCC 180
DB 4120 CAGTCTCCACGGGCTCAGATGTCACAGCTTCCTCCACCGAGGCGCACAGCCC 4179
QY 181 CTGGAGACCCCTTGAAGTGGCTTGTTCCTGTGATATCTCTTGGCCCTATAGTGG 240
DB 4180 CTGGAGACCCCTTGAAGTGGCTTGTTCCTGTGATATCTCTTGGCCCTATAGTGG 4239
QY 241 GCTTGGCCATTTGCCCTTCACTCCAGATCTCTCTTCCAGGTCAGGAGTGCATCTTGA 300
DB 4240 GCTTGGCCATTTGCCCTTCACTCCAGATCTCTCTTCCAGGTCAGGAGTGCATCTTGA 4299
QY 301 ACTTAATTTCCAGACCCCTTCACTTTCAGTTCCTAGTCTTACAGAGTGGACTTCTGATT 360
DB 4300 ACTTAATTTCCAGACCCCTTCACTTTCAGTTCCTAGTCTTACAGAGTGGACTTCTGATT 4359
QY 361 CTTTGTCTGTGCTGCTTGTAGCTCAGGTCAGGCTTAAGGCAAGTCTCTCACTGCTGC 420
DB 4360 CTTTGTCTGTGCTTGTAGCTCAGGTCAGGCTTAAGGCAAGTCTCTCACTGCTGC 4419
QY 421 CTGGGAGAGTCCAGAGCGCTGCAGGTCCTGTGGGTAGGATGCTGATGCCAGATT 480
DB 4420 CTGGGAGAGTCCAGAGCGCTGCAGGTCCTGTGGGTAGGATGCTGATGCCAGATT 4479
QY 481 TCCCGTTAGAGAGCCCTTCCCTATCTGACGCTCTAGCTTGTGTTACTTACTTGT 540
DB 4480 TCCCGTTAGAGAGCCCTTCCCTATCTGACGCTCTAGCTTGTGTTACTTACTTGT 4539
QY 541 CCACCTTAATTTCAAAATGATCCAGACACCAACAGCTGTGCACAGTTCTCTGGGGTTTCAGG 600
DB 4540 CCACCTTAATTTCAAAATGATCCAGACACCAACAGCTGTGCACAGTTCTCTGGGGTTTCAGG 4599

QY 601 AGGATGTAAGACATACCCCTTGCCCTTTCAGGCACCTATGCGCAGAGGGGGGCGAGTGACC 660
DB 4600 AGGATGTAAGACATACCCCTTGCCCTTTCAGGCACCTATGCGCAGAGGGGGGCGAGTGACC 4659
QY 661 TAGCAGAGGGCGGGAGCCAGCAGATGGATACACTCAGAGGAGCCTGCGAGCGGAGAG 720
DB 4660 TAGCAGAGGGCGGGAGCCAGCAGATGGATACACTCAGAGGAGCCTGCGAGCGGAGAG 4719
QY 721 GCAGAGGAGAGGAGGAGCTTACAGCTTCTGC 751
DB 4720 GCAGAGGAGAGGAGGAGCTTACAGCTTCTGC 4750
RESULT 2
AF064860 170121 bp DNA linear PRI 05-MAR-2002
LOCUS Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
DEFINITION sequence.
AF064860
ACCESSION AF064860.2 GI:18958624
VERSION HTG; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170121)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polly,A., Menzel,J., Delabar,J., Kumpf,K., Lehmann,R.,
Peterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudo,H., Kawasaki,K., Asakawa,S.,
Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
Antonakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Rieselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,
Borzum,K., Groliner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M. Laune.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
JOURNAL Nature 405 (6784), 311-319 (2000)
MEDLINE 20289799
PUBMED 10830953
REFERENCE 2 (bases 1 to 170121)
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 170121)
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 170121)
Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Feb 27, 2002 this sequence version replaced 91:3171153.
FEATURES
Location/Qualifiers
1. 170121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 70124"
BASE COUNT 45184 a 36756 c 37509 g 50672 t
ORIGIN

```
Query Match      100.0%; Score 751; DB 9; Length 170121;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAGCAGACTTACACTCAGGACTGCCAACACCCCTCCCGGGTTTCCCTGGTCTCGAAT 60
    |||
Db 85814 AAAGCAGACTTACACTCAGGACTGCCAACACCCCTCCCGGGTTTCCCTGGTCTCGAAT 85873
    |||

Qy 61 TCCATCCCTGGTTCCACCTGTATACATCACACTCCCTTCAAGACCAAGTGCAGATGC 120
    |||
Db 85874 TCCATCCCTGGTTCCACCTGTATACATCACACTCCCTTCAAGACCAAGTGCAGATGC 85933
    |||

Qy 121 CACGTCCTTCCAGGGGCTCAGAAATGCTCACAGCTTCCCTCCACCGAGGGCCACAGCC 180
    |||
Db 85934 CACGTCCTTCCAGGGGCTCAGAAATGCTCACAGCTTCCCTCCACCGAGGGCCACAGCC 85993
    |||

Qy 181 CTGAGACCCCTTGAAGTGAAGTCTTTCCTTGCATCTCTTTCGGGCTCATAGTGG 240
    |||
Db 85994 CTGAGACCCCTTGAAGTGAAGTCTTTCCTTGCATCTCTTTCGGGCTCATAGTGG 86053
    |||

Qy 241 GCTTGGCCATTGTCCTTCACTCCAGATCTCTCTTTCAGGTCAGGAAGTGCATCTGA 300
    |||
Db 86054 GCTTGGCCATTGTCCTTCACTCCAGATCTCTCTTTCAGGTCAGGAAGTGCATCTGA 86113
    |||

Qy 301 ACTTAACCTTCCAGACCCCTTCAAGTTCAGTTCCTTCCAGTCTTAGAGAGTGGACTTCGATT 360
    |||
Db 86114 ACTTAACCTTCCAGACCCCTTCAAGTTCAGTTCCTTCCAGTCTTAGAGAGTGGACTTCGATT 86173
    |||

Qy 361 CCTTGTCTCTGTGCTGTAGCTCAGGTCAGGCTTAAGCAAGTCTCCCTCACTCTGC 420
    |||
Db 86174 CCTTGTCTCTGTGCTGTAGCTCAGGTCAGGCTTAAGCAAGTCTCCCTCACTCTGC 86233
    |||

Qy 421 CTGGGAGAGTCCAGAGCCTCCAGTGCCTGTGCGGTAGATGCTGATCCCGAGATT 480
    |||
Db 86234 CTGGGAGAGTCCAGAGCCTCCAGTGCCTGTGCGGTAGATGCTGATCCCGAGATT 86293
    |||

Qy 481 TCCGTTTAGAGAGCTTTCCTATCTGAGGCTCTAGCTTTCTGTTACTTACTTGT 540
    |||
Db 86294 TCCGTTTAGAGAGCTTTCCTATCTGAGGCTCTAGCTTTCTGTTACTTACTTGT 86353
    |||

Qy 541 CCACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
    |||
Db 86354 CCACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86413
    |||

Qy 601 AGGATGTAAGACATACCCCTTCCCTTCCAGCCTATGCGCAGAGGGGGCGAGTGC 660
    |||
Db 86414 AGGATGTAAGACATACCCCTTCCCTTCCAGCCTATGCGCAGAGGGGGCGAGTGC 86473
    |||

Qy 661 TAGCAGAGGGGGGGAGCCAGCAGATGGATACACTCAGAGGAGCTGCGAGCAGGCAG 720
    |||
Db 86474 TAGCAGAGGGGGGGAGCCAGCAGATGGATACACTCAGAGGAGCTGCGAGCAGGCAG 86533
    |||

Qy 721 GCAGAGGAGAGGAGGCTTACAGTTCTGC 751
    |||
Db 86534 GCAGAGGAGAGGAGGCTTACAGTTCTGC 86564
    |||
```

```
RESULT 3
HS21C080      340000 bp      DNA      linear      PRI 24-MAY-2000
LOCUS      Homo sapiens chromosome 21 segment HS21C080.
DEFINITION      AL163280 AP001735 BA000005
ACCESSION      AL163280.2 GI:7717369
VERSION
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T.,
Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E.,
Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,
Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,
```

Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sakaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Brandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M. L.

Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GSF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8585, Japan,
* e.mail: sakaki@gsf.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: shimizu@db.med.keio.ac.jp
* URL: http://adenine.db.med.keio.ac.jp/
and
* GSF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Innestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.zz-berlin.mpg.de/.

Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
1. 125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P160L9, 5' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121897"
73410..24333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF064860"
224137..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121782"
286628..340000
/organism="Homo sapiens"


```

QY 421 CTGGGAGAGTCCAGGAGCTGCACGCTGCTGTGCGGTAGGATGCTGATGCCAGATT 480
Db 159642 CTGGGAGAGTCCAGGAGCTGCACGCTGCTGTGCGGTAGGATGCTGATGCCAGATT 159701
QY 481 TCCCGTTAGAGAGCCCTTCCCTATCCTCAGCGCTCTAGCTTTGTGTTACTTACTTGT 540
Db 159702 TCCCGTTAGAGAGCCCTTCCCTATCCTCAGCGCTCTAGCTTTGTGTTACTTACTTGT 159761
QY 541 CCACTTTAATTCAAATGTACCAACCAACAGCTTGTGCACAGTCTCTCTGGGTTTCAGG 600
Db 159762 CCACTTTAATTCAAATGTACCAACCAACAGCTTGTGCACAGTCTCTCTGGGTTTCAGG 159821
QY 601 AGGATGTAGACATACCCCTTGCCTTCAGGCACTATGCCAGCAAGGGGGCAGTGACC 660
Db 159822 AGGATGTAGACATACCCCTTGCCTTCAGGCACTATGCCAGCAAGGGGGCAGTGACC 159881
QY 661 TAGGCAGAGCGGGGAGCCAGCAGATGGGATACACTACAGAGAGCCTCAGCAGGCAGAG 720
Db 159882 TAGGCAGAGCGGGGAGCCAGCAGATGGGATACACTACAGAGAGCCTCAGCAGGCAGAG 159941
QY 721 GCAGAGAGAAGGGAGGTCTACAGTTCTGC 751
Db 159942 GCAGAGAGAAGGGAGGTCTACAGTTCTGC 159972

RESULT 4
AC073231 185982 bp DNA linear HTG 17-JUN-2000
LOCUS Homo sapiens chromosome 17 clone RP11-729A4, WORKING DRAFT
DEFINITION AC073231
SEQUENCE, 25 unordered pieces.
ACCESSION AC073231
VERSION 1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 185982)
Waterston,R.H.
Direct Submission
Submitted (10-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1724: contig of 1724 bp in length
* 1725 1824: gap of unknown length
* 1825 2971: contig of 1147 bp in length
* 2972 3072: gap of unknown length
* 3072 4660: contig of 1589 bp in length
* 4661 4761: gap of unknown length
* 4761 5930: contig of 1170 bp in length
* 5931 6031: gap of unknown length
* 6031 8127: contig of 2097 bp in length
* 8128 8228: gap of unknown length
* 8228 10160: contig of 1933 bp in length
* 10161 10261: gap of unknown length
* 10261 13236: contig of 2976 bp in length
* 13237 13337: gap of unknown length
* 13337 16603: contig of 3267 bp in length
* 16604 16703: gap of unknown length

20684: contig of 3981 bp in length
20784: gap of unknown length
24835: contig of 4051 bp in length
24935: gap of unknown length
29727: contig of 4792 bp in length
29827: gap of unknown length
34947: contig of 5120 bp in length
35047: gap of unknown length
41164: contig of 6117 bp in length
41264: gap of unknown length
47050: contig of 5786 bp in length
47151: gap of unknown length
53084: contig of 5934 bp in length
53184: gap of unknown length
60624: contig of 7440 bp in length
60724: gap of unknown length
69057: contig of 8333 bp in length
69157: gap of unknown length
78529: contig of 9372 bp in length
78629: gap of unknown length
85475: contig of 6846 bp in length
85575: gap of unknown length
97151: contig of 11576 bp in length
97251: gap of unknown length
113111: contig of 15860 bp in length
113211: gap of unknown length
129870: contig of 16659 bp in length
129970: gap of unknown length
145568: contig of 15598 bp in length
145668: gap of unknown length
184596: contig of 18928 bp in length
164696: gap of unknown length
164597 185982: contig of 21286 bp in length.

FEATURES
Location/Qualifiers
1..185982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-729A4"
1..1724
/note="assembly_name:Contig3"
1825..2971
/note="assembly_name:Contig4"
3072..4660
/note="assembly_name:Contig5"
4761..5930
/note="assembly_name:Contig6"
clone_end:T7
vector_side:right"
6031..8127
/note="assembly_name:Contig7"
8228..10160
/note="assembly_name:Contig8"
10261..13236
/note="assembly_name:Contig9"
13337..16603
/note="assembly_name:Contig10"
16704..20684
/note="assembly_name:Contig11"
20785..24835
/note="assembly_name:Contig12"
24936..29727
/note="assembly_name:Contig13"
clone_end:SP6
vector_side:right"
29828..34947
/note="assembly_name:Contig14"
35048..41164
/note="assembly_name:Contig15"
41265..47050
/note="assembly_name:Contig16"
47151..53084
/note="assembly_name:Contig17"

```



```
* 145965 147112: contig of 1148 bp in length
* 147113 147212: gap of unknown length
* 148514 148514: contig of 1302 bp in length
* 148515 148614: gap of unknown length
* 148615 148615: contig of 1224 bp in length
* 148938 148938: gap of unknown length
* 149339 151609: contig of 1671 bp in length
* 151610 151709: gap of unknown length
* 151710 153006: contig of 1297 bp in length
* 153007 153106: gap of unknown length
* 153107 154411: contig of 1335 bp in length
* 154642 154741: gap of unknown length
* 154742 156383: contig of 1642 bp in length
* 156384 156483: gap of unknown length
* 156484 157629: contig of 1146 bp in length
* 157630 157729: gap of unknown length
* 157730 159938: contig of 2209 bp in length
* 159939 160038: gap of unknown length
* 160039 161113: contig of 1075 bp in length
* 161114 161213: gap of unknown length
* 161214 163271: contig of 2058 bp in length
* 163272 163371: gap of unknown length
* 163372 164572: contig of 1201 bp in length
* 164573 164672: gap of unknown length
* 164673 166228: contig of 1656 bp in length
* 166329 166428: gap of unknown length
* 166429 168141: contig of 1713 bp in length
* 168142 168241: gap of unknown length
* 168242 169716: contig of 1475 bp in length
* 169717 169816: gap of unknown length
* 169817 171433: contig of 1827 bp in length
* 171644 171743: gap of unknown length
* 171744 173759: contig of 2016 bp in length
* 173760 173859: gap of unknown length
* 173860 175166: contig of 1307 bp in length
* 175167 175266: gap of unknown length
* 175267 177198: contig of 1932 bp in length
* 177199 177298: gap of unknown length
* 177299 179669: contig of 2371 bp in length
* 179670 179769: gap of unknown length
* 179770 180899: contig of 1130 bp in length
* 180900 180999: gap of unknown length
* 181000 182990: contig of 1991 bp in length
* 182991 183090: gap of unknown length
* 183091 184818: contig of 1728 bp in length
* 184819 184918: gap of unknown length
* 184919 186377: contig of 1459 bp in length
* 186378 186477: gap of unknown length
* 186478 187580: contig of 1103 bp in length
* 187581 189577: gap of unknown length
* 187582 189577: contig of 1897 bp in length
* 189578 189677: gap of unknown length
* 189678 191305: contig of 1628 bp in length
* 191306 191405: gap of unknown length
* 191406 194197: contig of 2792 bp in length
* 194198 194297: gap of unknown length
* 194298 197409: contig of 3112 bp in length
* 197410 197509: gap of unknown length
* 197510 201115: contig of 3606 bp in length
* 201116 201215: gap of unknown length
* 201216 203304: contig of 2069 bp in length
* 203305 203404: gap of unknown length
* 203405 206377: contig of 2973 bp in length
* 206378 206477: gap of unknown length
* 206478 209607: contig of 3130 bp in length
* 209608 209707: gap of unknown length
* 209708 213870: contig of 4163 bp in length
* 213871 213970: gap of unknown length
* 213971 215881: contig of 1911 bp in length
* 215882 215981: gap of unknown length
* 215982 219091: contig of 3110 bp in length
* 219092 219191: gap of unknown length
* 219192 224729: contig of 5538 bp in length
```

```
* 224730 224829: gap of unknown length
* 224830 229823: contig of 4994 bp in length
* 229824 229923: gap of unknown length
* 229924 235164: contig of 5241 bp in length
* 235165 235264: gap of unknown length
* 235265 244581: contig of 9317 bp in length
* 244582 244681: gap of unknown length
* 244682 251899: contig of 7218 bp in length
FEATURES
    source
        1..251899
            Location/Qualifiers
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="UNK"
                /clone="Rp24-296C19"
        1..30881
            /note="assembly_name:Contig100"
        30982..75783
            /note="assembly_name:Contig101"
        75884..131780
            /note="assembly_name:Contig102"
        131881..132912
            /note="assembly_name:Contig21"
        133013..134220
            /note="assembly_name:Contig40"
        134321..135773
            /note="assembly_name:Contig46"
        135874..137463
            /note="assembly_name:Contig47"
        137564..138714
            /note="assembly_name:Contig48"
        138815..140184
            /note="assembly_name:Contig50"
        140285..141523
            /note="assembly_name:Contig51"
        141624..143237
            /note="assembly_name:Contig52"
        143338..144648
            /note="assembly_name:Contig54"
        144749..145864
            /note="assembly_name:Contig55"
        145965..147112
            /note="assembly_name:Contig56"
        147213..148514
            /note="assembly_name:Contig57"
        148615..149838
            /note="assembly_name:Contig59"
        149939..151609
            /note="assembly_name:Contig60"
        151710..153006
            /note="assembly_name:Contig63"
        153107..154641
            /note="assembly_name:Contig63"
```

Query Match 3 24; Score 24; DB 2; Length 251899;
Best Local Similarity 100.0%; Pred.No. 0.02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 714 GGCAGAGCAGAGGAGGAGGAGG 737
|||||
Db 181678 GGCAGAGCAGAGGAGGAGGAGG 181701

RESULT 6

AC007237/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1

AC007237 Homo sapiens PAC clone RP5-911M23 from 7, complete sequence.
AC007237 Homo sapiens PAC clone RP5-911M23 from 7, complete sequence.
AC007237.3 GI:5708494
HTG.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 130815)

AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
---------	-------	---------	---------	--------	-----------	---------	-------	---------	-----------	---------	-------	---------	---------

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequencing of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:edgreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 5:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>) or from Pieter de Jong.

REGION: PCFRA2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-144P23, 200 base overlap;
the clone sequenced to the right is RP4-731H19, 200 base overlap.
Actual start of this clone is at base position 197 of RP3-911M23.

FEATURES	actual	end is at base position 42233 of RP4-731H19.
source	Location/Qualifiers 1. .130815	
	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7"	
repeat_region	/clone="RP5-911M23" /clone_lib="RPC1-5" 2670. .2749	
repeat_region	/rpt_family="MIR" 2938. .3229	
repeat_region	/rpt_family="Alu" 3263. .3354	
repeat_region	/rpt_family="MIR" 6094. .6131	
repeat_region	/rpt_family="AT_rich" 6980. .7135	
repeat_region	/rpt_family="L2" 8713. .9066	
repeat_region	/rpt_family="MALR" 9086. .9111	
repeat_region	/rpt_family="AT_rich" 9228. .9575	
repeat_region	/rpt_family="L1" 9897. .10209	
repeat_region	/rpt_family="L1" 11023. .11339	
repeat_region	/rpt_family="L1" 11368. .11946	
repeat_region	/rpt_family="L1" 11953. .12436	
repeat_region	/rpt_family="L1" 12437. .12816	
repeat_region	/rpt_family="MALR" 12817. .14171	
repeat_region	/rpt_family="L1" 14172. .14217	
repeat_region	/rpt_family="(CA)n" 14312. .14507	
repeat_region	/rpt_family="L1" 16108. .16187	
repeat_region	/rpt_family="MIR" 16984. .17045	
repeat_region	/rpt_family="(CATA)n" 17678. .17809	
repeat_region	/rpt_family="MER1_type" 17809. .18034	
misc_feature	/note="similar to EST W92351 (NID:gl424735) ze15e06.rl"	
misc_feature	/note="similar to Mus musculus EST W84941 (NID:g1538565) m42b04.rl"	
misc_feature	/note="similar to Mus musculus EST AA230716 (NID:g1853064) mw07b10.rl"	
misc_feature	/note="similar to EST W52136 (NID:gl349306) zc48g08.rl"	
misc_feature	/note="similar to EST H65927 (NID:g1024667) yr69g02.rl"	
misc_feature	/note="similar to EST H41039 (NID:g917091) yn92dl1.rl"	
misc_feature	/note="similar to Mus musculus EST AA102890 (NID:g1649066) mo07g08.rl"	
misc_feature	/note="similar to Rattus norvegicus EST AI547734 (NID:g4465222)"	
misc_feature	/note="similar to EST AI198351 (NID:g3750957) q162c01.xl"	
misc_feature	/note="similar to Mus musculus EST A1645993 (NID:g4724468) vv53h09.xl"	
misc_feature	17955. .18309	


```

/note="similar to EST H65927 (NID:g1024667) yr69g02.rl"
17978..18688
/note="similar to EST A1140111 (NID:g3647568) qa95c03.x1"
17990..18152
/note="similar to EST A1523891 (NID:g4438026) tg97f10.x1"
18008..18700
/note="similar to EST A1189286 (NID:g3740495) qd05c03.x1"
18016..18700
/note="similar to Rattus sp. EST AA849740 (NID:g2937280)"
18094..18551
/note="similar to EST AA025474 (NID:g1490938) ze90a07.rl"
18113..18684
/note="similar to Mus musculus EST AA930956 (NID:g3079389)
yr79g02.rl"
18121..18706
/note="similar to EST A1032877 (NID:g3253574) owl3g04.x1"
18138..18624
/note="similar to EST W37750 (NID:g1319344) zcl2b11.rl"
18158..18711
/note="similar to Mus musculus EST A1132679 (NID:g3602695)
ue3gd10.x1"
18163..18688
/note="similar to EST W30822 (NID:g1311966) ze65g11.rl"
18188..18704
/note="similar to EST A1190417 (NID:g3741626) qd53e01.x1"
18188..18688
/note="similar to EST N36122 (NID:g1157264) yy32g07.s1"
18196..18711
/note="similar to EST AA889268 (NID:g3016147) ak19b01.s1"
18209..18698
/note="similar to EST A1066676 (NID:g3366962) oz87c09.x1"
18238..18688
/note="similar to EST H20171 (NID:g8888866) yn61a06.s1"
18243..18700
/note="similar to EST A1523891 (NID:g4438026) tg97f10.x1"
18246..18705
/note="similar to EST A1016758 (NID:g3231094) ov27c10.x1"
18273..18693
/note="similar to EST D19615 (NID:g499931)"
18294..18711
/note="similar to EST R50840 (NID:g812742) yg62e04.s1"
18316..18710
/note="similar to EST A1752939 (NID:g5131203) cr02e11.x2"
18330..18711

```

Query Match 2.9%; Score 22; DB 9; Length 130815;

Best Local Similarity 100.0%; Pred. No. 0.33;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 AATGCTCACCAGCTTCCTCTCC 163

Db 4152 AATGCTCACCAGCTTCCTCTCC 4131

RESULT 7
AC105371/c
LOCUS
DEFINITION Sus scrofa clone RP44-44P5, WORKING DRAFT SEQUENCE, 6 ordered
pieces.
ACCESSION AC105371
VERSION AC105371.2 GI:21629287
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactylia; Suina; Suidae; Sus.
1 (bases 1 to 133526)
Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Lario, P.,
Lee-lin, S.-Q., Legaspi, R., Maduro, O.L., Maduro, V.B.,
Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D.,

McCluskey, J.C., McDowell, J., Paguirigan, C., Pearson, R.,
Portnoy, M.E., Prasad, A., Schueller, M.G., Stantropop, S., Thomas, J.W.,
Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 133526)
Green, E.D.
Direct Submission
Submitted (03-JAN-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 133526)
Green, E.D.
Direct Submission
Submitted (29-JUN-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jun 29, 2002 this sequence version replaced gi:18042290.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoehngri@nisc.nih.gov

----- Project Information

Center project name: cda

Center clone name: 044P05

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 132111 bases at least Q40

Consensus quality: 132642 bases at least Q30

Consensus quality: 132906 bases at least Q20

Insert size: 131000; agarose-tp

Quality coverage: 10.86x in Q20 bases; agarose-tp

Quality coverage: 10.70x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 9713: contig of 9713 bp in length

* 9714 9813: gap of unknown length

* 9814 23987: contig of 14174 bp in length

* 23988 24087: gap of unknown length

* 24088 27877: contig of 3790 bp in length

* 27878 27977: gap of unknown length

* 27978 32431: contig of 4454 bp in length

* 32432 32531: gap of unknown length

* 32532 121537: contig of 93006 bp in length

* 121538 121637: gap of unknown length

* 121638 133526: contig of 11889 bp in length.

Location/Qualifiers

1..133526

/organism="Sus scrofa"

FEATURES
source

```

/db_xref="taxon:9823"
/clone="RP44-44P5"
/clone_lib="RP44"
1. 56855
/note="clone overlaps with GenBank Accession Number
AC105367 clone RP44-123J14 (center project name cdb)"
1. 9713
/note="assembly_fragment
clone_end:SP6
vector_side:left"
9814. 23987
/note="assembly_fragment"
24088. 27877
/note="assembly_fragment"
27978. 32431
/note="assembly_fragment"
32532. 121537
/note="assembly_fragment"
121638. 133526
/note="assembly_fragment
clone_end:T7
vector_side:right"
39767 a 26117 c 26975 g 40167 t 500 others
BASE COUNT
ORIGIN

Query Match 2.9% Score 22; DB 2; Length 133526;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 GTCACACAGCTTCCTCCACC 156
|||||
Db 78301 GTCACACAGCTTCCTCCACC 78280

RESULT 8
AC125568
LOCUS
DEFINITION
Rattus norvegicus clone CH230-919, *** SEQUENCING IN PROGRESS ***,
70 unordered pieces.
ACCESSION
AC125568
VERSION
AC125568.1 GI:21628981
KEYWORDS
HTG: HTGS-PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 146487)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chakrabarti,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.B., Cox,C., Coyle,M.B., Datchakraborty,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loughsegh,H.,
Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapue,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,2., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,H.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 146487)
Worley,K.C.
Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 146487)
Worley,K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDEE
Center clone name: CH230-919
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 75912 bases at least Q40
Consensus quality: 80934 bases at least Q30
Consensus quality: 84235 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1214: contig of 1214 bp in length
* 1215: gap of unknown length
* 1315: contig of 1023 bp in length
* 2338: gap of unknown length
* 2438: contig of 1269 bp in length
* 3706: gap of unknown length
* 3806: gap of unknown length
* 5421: contig of 1615 bp in length
* 5521: gap of unknown length
* 7226: contig of 1705 bp in length
* 7326: gap of unknown length
* 7727: contig of 1401 bp in length
* 8728: gap of unknown length
* 8828: contig of 1050 bp in length
* 9877: gap of unknown length
* 11300: contig of 1323 bp in length
* 11400: gap of unknown length
* 11401: contig of 1578 bp in length
* 13078: gap of unknown length

```

* 13079 14470: contig of 1392 bp in length
 * 14471 14570: gap of unknown length
 * 14571 15835: contig of 1265 bp in length
 * 15835 15935: gap of unknown length
 * 15935 17064: contig of 1129 bp in length
 * 17064 17164: gap of unknown length
 * 17164 18438: contig of 1274 bp in length
 * 18438 18538: gap of unknown length
 * 18538 19759: contig of 1221 bp in length
 * 19759 19859: gap of unknown length
 * 19859 21187: contig of 1328 bp in length
 * 21187 21287: gap of unknown length
 * 21287 23037: contig of 1750 bp in length
 * 23037 23137: gap of unknown length
 * 23137 24195: contig of 1058 bp in length
 * 24195 24295: gap of unknown length
 * 24295 25911: contig of 1516 bp in length
 * 25911 25912: gap of unknown length
 * 25912 27650: contig of 1739 bp in length
 * 27650 27750: gap of unknown length
 * 27750 29440: contig of 1690 bp in length
 * 29440 29540: gap of unknown length
 * 29540 30879: contig of 1339 bp in length
 * 30879 30979: gap of unknown length
 * 30979 32394: contig of 1415 bp in length
 * 32394 32494: gap of unknown length
 * 32494 34067: contig of 1573 bp in length
 * 34067 34167: gap of unknown length
 * 34167 35410: contig of 1243 bp in length
 * 35410 35510: gap of unknown length
 * 35510 36916: contig of 1306 bp in length
 * 36916 38030: contig of 1114 bp in length
 * 38030 38130: gap of unknown length
 * 38130 39608: contig of 1478 bp in length
 * 39608 39708: gap of unknown length
 * 39708 41329: contig of 1621 bp in length
 * 41329 41429: gap of unknown length
 * 41429 43399: contig of 1970 bp in length
 * 43399 43499: gap of unknown length
 * 43499 45424: contig of 1925 bp in length
 * 45424 45524: gap of unknown length
 * 45524 47337: contig of 1713 bp in length
 * 47337 47338: gap of unknown length
 * 47338 50200: contig of 2863 bp in length
 * 50200 50300: gap of unknown length
 * 50300 51695: contig of 1395 bp in length
 * 51695 51795: gap of unknown length
 * 51795 53166: contig of 1371 bp in length
 * 53166 53266: gap of unknown length
 * 53266 55360: contig of 1994 bp in length
 * 55360 55361: gap of unknown length
 * 55361 56638: contig of 1278 bp in length
 * 56638 56738: gap of unknown length
 * 56738 58995: contig of 2257 bp in length
 * 58995 59095: gap of unknown length
 * 59095 60222: contig of 1127 bp in length
 * 60222 60322: gap of unknown length
 * 60322 61599: contig of 1277 bp in length
 * 61599 61699: gap of unknown length
 * 61699 63238: contig of 1539 bp in length
 * 63238 63338: gap of unknown length
 * 63338 64893: contig of 1555 bp in length
 * 64893 64993: gap of unknown length
 * 64993 66140: contig of 1147 bp in length
 * 66140 66240: gap of unknown length
 * 66240 69520: contig of 3280 bp in length
 * 69520 69620: gap of unknown length
 * 69620 71388: contig of 1668 bp in length
 * 71388 71389: gap of unknown length
 * 71389 72577: contig of 1189 bp in length
 * 72577 72678: gap of unknown length
 * 72678 74257: contig of 1580 bp in length

* 74258 74357: gap of unknown length
 * 74357 76629: contig of 2272 bp in length
 * 76629 76630: gap of unknown length
 * 76630 79513: contig of 2784 bp in length
 * 79513 79613: gap of unknown length
 * 79613 82612: contig of 2999 bp in length
 * 82612 82712: gap of unknown length
 * 82712 85168: contig of 2456 bp in length
 * 85168 85268: gap of unknown length
 * 85268 86830: contig of 1562 bp in length
 * 86830 86931: gap of unknown length
 * 86931 89131: contig of 2201 bp in length
 * 89131 89132: gap of unknown length
 * 89132 89232: contig of 2182 bp in length
 * 89232 91414: gap of unknown length
 * 91414 94288: contig of 2775 bp in length
 * 94288 94388: gap of unknown length

Query Match 2.9%; Score 22; DB 2: Length 146487;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 706 CTGCAGCAGCAGCAGCAGCAGG 727
 |||||
 DB 46224 CTGCAGCAGCAGCAGCAGCAGG 46245

RESULT 9
 AL356740/c

LOCUS AL356740
 DEFINITION Human DNA sequence from clone RP11-120K24 on chromosome 13,
 complete sequence.

ACCESSION AL356740
 VERSION AL356740.19 GI:16972886

KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 170928)
 TITLE Direct Submission

AUTHORS Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

JOURNAL On Nov 16, 2001 this sequence version replaced gi:13751403.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

COMMENT This sequence was finished as follows unless otherwise noted; all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-120K24 is from the library RPCR-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6
 This sequence is the entire insert of clone RP11-120K24 The true

right end of clone RP11-73M11 is at 30298 in this sequence.

FEATURES
 source
 1. .170928
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-120K24"
 /clone_lib="RPCI-11.1"
 /notes="Sequence from overlapping clone RP11-73M11
 17045. .17099
 (AL356752).
 restriction digest."
 28286. .28355
 /note="Sequence from overlapping clone RP11-73M11
 (AL356752). Assembly confirmed by
 restriction digest."
 46727. .46797
 /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 57005. .57302
 /note="Sequence from uni-directional dGTP big dye
 terminator reads only."
 72039
 /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 550bp by restriction digest data."
 106842
 /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 1.3kb by restriction digest data."
 37569 a 44187 c 47897 g 41275 t

BASE COUNT 37569 a 44187 c 47897 g 41275 t

ORIGIN
 Query Match 2.9%; Score 22; DB 9; Length 170928;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 CAGAGGAGCTGCAGGAGGAG 718
 |||||
 Db 45553 CAGAGGAGCTGCAGGAGGAG 45532

RESULT 10
 AL731778/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-475B13 on chromosome 2, complete
 sequence.
 ACCESSION AL731778
 VERSION AL731778.12 GI:21272461
 KEYWORDS HTG.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, M.
 Direct Submission
 Submitted (17-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Jul 10, 2002 this sequence version replaced gi:21690013.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW:
 SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6.

FEATURES
 Location/Qualifiers
 source
 1. .211430
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-475B13"
 /clone_lib="RPCI-23"
 53376 a 52028 c 52641 g 53385 t

BASE COUNT 53376 a 52028 c 52641 g 53385 t

ORIGIN
 Query Match 2.9%; Score 22; DB 10; Length 211430;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 GCAGGCGAGGAGGAGGAGGAG 736
 |||||
 Db 204810 GCAGGCGAGGAGGAGGAGGAG 204789

RESULT 11
 AL137249
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-111L24 on chromosome
 13q31.3-32.3, complete sequence.
 ACCESSION AL137249
 VERSION AL137249.29 GI:18655955
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Garner, P.
 Direct Submission
 Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Feb 13, 2002 this sequence version replaced gi:18121451.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW:
 SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>

RP11-111L24 is from the library RPCI-11.1 constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECT08: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-111L24. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone RP11-295B17 is at 104579 in this sequence. The true right end of clone RP11-261P24 is at 2000 in this sequence.

FEATURES
 source
 1..106578
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="q31.3-32.3"
 /clone="RP11-111L24"
 /clone_lib="RPCI-11.1"
 79533
 misc_feature
 /note="Tandem repeat. Forced join. Gap size estimated to be approximately 800bp by restriction digest data"
 BASE COUNT 28669 a 25979 c 25140 g 26790 t
 ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 106578;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 CACCTCCCTTCAGGACCAG 111
 |||||||
 DB 65874 CACCTCCCTTCAGGACCAG 65894

RESULT 12

AC119727.0/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC119727 Accession AC119727

Fragment Name	Begin	End
AC119727_0	1	110000
AC119727_1	100001	210000
AC119727_2	200001	310000
AC119727_3	300001	369035

LOCUS AC119727 369035 bp DNA linear HTG 09-MAY-2002
 DEFINITION Homo sapiens chromosome 3 clone RP11-248D6, *** SEQUENCING IN
 PROGRESS ***, 28 unordered pieces.

AC119727

AC119727.2 GI:20452978

KEYWORDS HTG; HTGS_PHASE1.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 369035)

REFERENCE
 AUTHORS
 Murny,D.M., Adams,C., Adio-Oduola,B., Ali-olsman,F.R., Allen,C., Alsbrooks,S.L., Amarutunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobsen,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,D., Vera,V., Villalobos,B., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

```

* 15571 18734: contig of 3164 bp in length
* 18735 18834: gap of unknown length
* 18835 23420: contig of 6586 bp in length
* 23421 25220: gap of unknown length
* 25221 32590: contig of 6970 bp in length
* 32591 40944: contig of 8354 bp in length
* 40945 41044: gap of unknown length
* 41045 47022: contig of 5978 bp in length
* 47023 47122: gap of unknown length
* 47123 54309: contig of 7787 bp in length
* 54310 55009: gap of unknown length
* 55010 62601: contig of 7592 bp in length
* 62602 62701: gap of unknown length
* 62702 69505: contig of 6804 bp in length
* 69506 75688: gap of unknown length
* 75689 75688: gap of unknown length
* 75689 86668: contig of 10980 bp in length
* 86669 86768: gap of unknown length
* 86769 95886: contig of 12918 bp in length
* 95887 99786: gap of unknown length
* 99787 107918: contig of 8132 bp in length
* 107919 108018: gap of unknown length
* 108019 119656: contig of 11638 bp in length
* 119657 119756: gap of unknown length
* 119757 137481: contig of 17725 bp in length
* 137482 137581: gap of unknown length
* 137582 153929: contig of 16348 bp in length
* 153930 154029: gap of unknown length
* 154030 167945: contig of 13916 bp in length
* 167946 168045: gap of unknown length
* 168046 184922: contig of 16877 bp in length
* 184923 217509: contig of 32487 bp in length
* 217510 217609: gap of unknown length
* 217610 250039: contig of 32430 bp in length
* 250040 250139: gap of unknown length
* 250140 294258: contig of 44119 bp in length
* 294259 294358: gap of unknown length
* 294359 329184: contig of 34826 bp in length
* 329185 329284: gap of unknown length
* 329285 369035: contig of 39751 bp in length.

```

FEATURES

```

Location/Qualifiers
1..369035
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"

```

BASE COUNT 109667 a 75630 c 74833 g 106175 t 2730 others

ORIGIN

```

Query Match 2.8% Score 21; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 730 AAGGAGGCTACACGTTCTG 750

Db 13853 AAGGAGGCTACACGTTCTG 13833

RESULT 13

```

AC106847/c 135431 bp DNA linear PRI 02-MAR-2002
LOCUS AC106847 Homo sapiens chromosome 8, clone CTD-2333A6, complete sequence.
DEFINITION AC106847
ACCESSION AC106847
VERSION AC106847.3 GI:19071668
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135431)

```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Homo sapiens chromosome 8, clone CTD-2333A6
Unpublished
2 (bases 1 to 135431)
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (12-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 135431)
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamarez,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 135431)
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamarez,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.
 Direct Submission
 Submitted (02-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 2, 2002 this sequence version replaced gi:18643660.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23195
 Center clone name: 2333_A_6

FEATURES

source

Location/Qualifiers

```

1..135431
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="8"
   /map="8"
   /clone="CTD-2333A6"
344..1372
   /clone_lib="CIT01 Human BAC"
   /rpt_family="LTR5"
   complement(1697..1910)
   /rpt_family="MIR"
2305..2614
   /rpt_family="AluSg"
3747..3845
   /rpt_family="LIM4"
3855..3919
   /rpt_family="(TTG)n"
   complement(3922..4148)
   /rpt_family="LIM4"
4149..4177
   /rpt_family="(TTG)n"
   complement(4178..4216)
   /rpt_family="LIM4"
   complement(4527..4712)
   /rpt_family="MLT1J2"
   complement(4721..4850)
   /rpt_family="LIM4"
   complement(4921..5025)
   /rpt_family="MIR3"
   complement(5124..6503)
   /rpt_family="PTR5"
7796..8482
   /rpt_family="L2"
   complement(8483..8783)
   /rpt_family="AluSq"
8784..8967
   /rpt_family="L2"
   complement(9100..9179)
   /rpt_family="MER5A"
9190..9348
   /rpt_family="L2"
   complement(9360..9597)
   /rpt_family="AluJo"
9667..9703
   /rpt_family="(TC)n"
10350..10640
   /rpt_family="AluY"
   complement(10875..11181)
   /rpt_family="AluSc"
11491..11550
   /rpt_family="MIR"

```

```

repeat_region complement(11556..12157)
               /rpt_family="ERV1"
repeat_region 12174..12297
               /rpt_family="MIR"
repeat_region complement(12387..12455)
               /rpt_family="L2"
repeat_region 12976..13418
               /rpt_family="L2"
repeat_region 14650..14897
               /rpt_family="MIR"
repeat_region 14981..15277
               /rpt_family="AluJb"
repeat_region complement(16596..16826)
               /rpt_family="MER20"
repeat_region 16856..17039
               /rpt_family="MIR"
repeat_region complement(17043..17344)
               /rpt_family="AluSc"
repeat_region 17597..17656
               /rpt_family="MIR"
repeat_region complement(17662..18210)
               /rpt_family="ERV1"
repeat_region 18211..18403
               /rpt_family="MIR"
repeat_region complement(18493..18602)
               /rpt_family="L2"
repeat_region 18694..18803
               /rpt_family="(CCTG)n"
repeat_region 19082..19524
               /rpt_family="L2"
repeat_region 21649..21896
               /rpt_family="MIR"
repeat_region 21980..22285
               /rpt_family="AluJb"
repeat_region complement(23581..23807)
               /rpt_family="MER20"
repeat_region 23837..24020
               /rpt_family="MIR"
repeat_region complement(24024..24328)
               /rpt_family="AluSx"
repeat_region complement(25436..25639)
               /rpt_family="L2"
repeat_region complement(25669..25875)
               /rpt_family="AluSx"

```

Query Match 2.8% Score 21; DB 9; Length 135431;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 730 AAGGAGGTCTACAGTTCTG 750
 |||||
 Db 119920 AAGGAGGTCTACAGTTCTG 119900

RESULT 14
 AC128939/c

LOCUS
 DEFINITION

***, 45 unordered pieces.

AC128939

AC128939.1 GI:21953633

HTG: HTGS_PHASE1.

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 136417)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,

Barbours, S., L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbours, J., Benton, J., Bimage, K., Blankenburg, K., Bonnino, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

REFERENCE

AUTHORS

136417 bp DNA linear HTG 24-JUL-2002
 Rattus norvegicus clone CH230-434L13, *** SEQUENCING IN PROGRESS

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, T., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korval, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 2 (bases 1 to 136417)
 Worley, K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: KBEL
 Center clone name: CH230-434L13

 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 90451 bases at least Q40
 Consensus quality: 98862 bases at least Q30
 Consensus quality: 104856 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1033: contig of 1080 bp in length
 * 1081: gap of unknown length
 * 1181: contig of 1165 bp in length
 * 2346: gap of unknown length


```
repeat_region /rpt_family="AluJ" complement(2489..2618)
repeat_region /rpt_family="AluJ" complement(2620..2904)
repeat_region /rpt_family="AluJo" complement(3008..3176)
repeat_region /rpt_family="Char1188" complement(3194..3348)
repeat_region /rpt_family="L2" complement(3349..3658)
repeat_region /rpt_family="AluY" complement(3694..3852)
repeat_region /rpt_family="L2" complement(3853..3983)
repeat_region /rpt_family="L1MB3" complement(3984..4364)
repeat_region /rpt_family="L2" complement(4828..4918)
repeat_region /rpt_family="MLT1B" complement(4919..5220)
repeat_region /rpt_family="AluCo" complement(5221..5524)
repeat_region /rpt_family="MLT1B" complement(5719..5757)
repeat_region /rpt_family="L2" complement(6027..6242)
repeat_region /rpt_family="AluSx" complement(6243..6598)
repeat_region /rpt_family="THE1B" complement(6726..7202)
repeat_region /rpt_family="MLT1E" complement(7203..7239)
repeat_region /rpt_family="(TG)n" complement(7240..7527)
repeat_region /rpt_family="AluJo" complement(7528..7592)
repeat_region /rpt_family="MLT1E" complement(7734..7790)
repeat_region /rpt_family="MIR3" complement(7791..7833)
repeat_region /rpt_family="(TCCC)n" complement(7837..8139)
repeat_region /rpt_family="AluSg" complement(8141..8458)
repeat_region /rpt_family="AluSx" complement(8461..8657)
repeat_region /rpt_family="L1M1" complement(8658..8970)
repeat_region /rpt_family="AluSx" complement(8699..8705)
repeat_region /note="<30 qual SNG1 region" complement(8971..9211)
repeat_region /rpt_family="L1M1" complement(9212..9520)
repeat_region /rpt_family="AluSx" complement(9521..10126)
repeat_region /rpt_family="L1M1" complement(10120..11262)
repeat_region /rpt_family="L1M1" complement(11263..11375)
repeat_region /rpt_family="AluJb" complement(11420..11467)
repeat_region /rpt_family="AT-rich" complement(12098..12241)
repeat_region /rpt_family="AluSg" complement(12242..12536)
repeat_region /rpt_family="AluSg" complement(12537..12699)
repeat_region /rpt_family="AluSg" complement(13718..13777)
repeat_region /rpt_family="MIR" complement(13829..14044)
repeat_region /rpt_family="MIR"
```

```
repeat_region complement(14051..14499)
repeat_region /rpt_family="MLT1F" complement(14883..14940)
repeat_region /rpt_family="AluSg" complement(14953..15212)
repeat_region /rpt_family="AluJb" complement(15238..15302)
repeat_region /rpt_family="L1MB1" complement(15303..15476)
repeat_region /rpt_family="AluSx" complement(15477..15501)
repeat_region /rpt_family="(TA)n"
```

Query Match 2.8% Score 21: DB 9: Length 150507;
Best Local Similarity 100.08; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 730 AAGGGAGGTCTACAGTTCTG 750
|||||
DB 31059 AAGGGAGGTCTACAGTTCTG 31079

Search completed: April 12, 2003, 08:05:44
Job time : 2404.57 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:38:14 ; Search time 123.291 Seconds
(without alignments)
13717.504 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751

Sequence: 1 aaagcagactctagcactca.....ggaggtctacacgttctgc 751

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	100.0	10562	21	AAA93876 Human beta3Gal-T5
2	145	19.3	1745	23	AAS72914 DNA encoding novel
C 3	22	2.9	256	24	ABN24174 Human ORFX polynuc
C 4	20	2.7	903	23	ABL18235 Drosophila melanog
C 5	20	2.7	1347	20	AAAX31841 Human T-Box polype
C 6	20	2.7	2728	18	AAT58853 cDNA encoding high
C 7	20	2.7	2728	19	AAV28861 Rat prolina transp
C 8	20	2.7	3129	23	ABL18234 Drosophila melanog
C 9	19	2.5	65	24	ABN58123 Mouse spliced tran

C 10	18	2.4	298	21	AAA00717 Human colon cancer
C 11	18	2.4	438	18	AAT70809 Mouse anti-idiotyp
C 12	18	2.4	588	23	AAS72980 DNA encoding novel
C 13	18	2.4	1326	21	AAC43758 Zea mays DNA fragm
C 14	18	2.4	1427	23	AAS75162 DNA encoding novel
C 15	18	2.4	1572	21	AAS59691 Human secreted pro
C 16	18	2.4	2520	23	AAS68697 DNA encoding novel
C 17	18	2.4	2681	23	ABX42350 Genomic sequence #
C 18	18	2.4	2904	23	AAS72979 DNA encoding novel
C 19	18	2.4	3613	22	AAK80972 Human immune/haema
C 20	18	2.4	3621	22	AAK80975 Human immune/haema
C 21	18	2.4	3623	22	AAK80973 Human immune/haema
C 22	18	2.4	3767	22	AAK89269 Human digestive sy
C 23	18	2.4	3778	22	AAK89266 Human digestive sy
C 24	18	2.4	4231	22	AAK89267 Human digestive sy
C 25	18	2.4	6386	20	AAK99364 Maize shrunken-1 (
C 26	18	2.4	6402	22	AAK89268 Human digestive sy
C 27	18	2.4	7806	23	AAK88354 DNA encoding novel
C 28	18	2.4	10248	22	AAK81853 Human immune/haema
C 29	18	2.4	10248	23	ABK42351 Genomic sequence #
C 30	18	2.4	13309	22	AAK83980 Human immune/haema
C 31	18	2.4	27541	22	AAD17185 Streptomyces nous
C 32	18	2.4	122888	24	ABK83569 Streptomyces nous
C 33	18	2.4	125401	22	AAD17186 Novel human polynu
C 34	17	2.3	374	22	AAF67509 Novel human polynu
C 35	17	2.3	381	22	AAF67510 Novel human polynu
C 36	17	2.3	399	22	AAF64930 Human foetal liver
C 37	17	2.3	450	22	ABA52050 Human foetal liver
C 38	17	2.3	450	22	ABA52051 Human foetal liver
C 39	17	2.3	450	22	ABA21864 Probe #330 for gen
C 40	17	2.3	450	22	ABA21865 Probe #331 for gen
C 41	17	2.3	450	22	AAK00333 Human brain expres
C 42	17	2.3	450	22	AAK00334 Human brain expres
C 43	17	2.3	450	22	AAK25776 Human bone marrow
C 44	17	2.3	450	22	AAK25777 Human bone marrow
C 45	17	2.3	450	22	AA110405 Probe #338 for gen

ALIGNMENTS

RESULT 1

AAA93876

ID AAA93876 standard; DNA; 10562 BP.

XX

AC AAA93876;

XX

DT 15-JAN-2001 (first entry)

XX

DE Human beta3Gal-T5 encoding DNA.

XX

DE Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;

KW digestive system; beta3Gal-T5; ds.

XX

OS Homo sapiens.

XX

PN WO2000050608-A1.

XX

PD 31-AUG-2000.

XX

PF 24-FEB-2000; 2000WO-JP01070.

XX

PR 25-FEB-1999; 99JP-0047571.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX

PI Narimatsu H, Ishiki S, Togayachi A, Sasaki K;

XX

DR WPI; 2000-549409/50.

XX

PT Beta-1,3 galactose transferase and DNA encoding it, useful for

PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of

PT digestive system cancer

XX Claim 31: Page 103-111; 123pp; Japanese.
 PS
 XX This invention relates to a polypeptide (I) with beta-1,3 galactose
 CC transferase activity, or variants of (I) comprising amino acid additions,
 CC deletions and/or substitutions. Included in the invention is DNA encoding
 CC all or part of (I); expression vectors containing the DNA, host cells
 CC transformed by the vectors; a method for the preparation of the
 CC polypeptide by culture of the transformants or by expression in the milk
 CC of a transgenic mammal, and antibodies recognising (I). The beta-1,3
 CC galactose transferase protein transfers galactose by beta-1,3 bonding to
 CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
 CC GlcNAc[beta1-3Gal[beta1-4Glc]) to give Galbeta1-3GlcNAc. The protein and
 CC DNA encoding it are useful for the treatment and diagnosis of cancer of
 CC the digestive system. The present sequence represents a Beta3gal-T5
 CC encoding DNA sequence.
 XX
 SQ Sequence 10562 BP: 2610 A; 2415 C; 2574 G; 2963 T; 0 other;
 Query Match 100.0%; Score 751; DB 21; Length 10562;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 751: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAGCAGACTCTAGCACTCAGCAGCTGCCAACACCCCTCCCGGGTTCTCTGGTCTCGAAT 60
 DB 4000 AAAGCAGACTCTAGCACTCAGCAGCTGCCAACACCCCTCCCGGGTTCTCTGGTCTCGAAT 4059
 QY 61 TCCCATCCCGCTGGTTCCACCTGTTACATCAGACCTCCCGTTCAAGGACCAGTCAGATGC 120
 DB 4060 TCCCATCCCGCTGGTTCCACCTGTTACATCAGACCTCCCGTTCAAGGACCAGTCAGATGC 4119
 QY 121 CAGCTCCTTCACGGGGCTCAGAACTGCTCACCAGTTCCTCTCCACCGAGGCCACAGCCCC 180
 DB 4120 CAGCTCCTTCACGGGGCTCAGAACTGCTCACCAGTTCCTCTCCACCGAGGCCACAGCCCC 4179
 QY 181 CTGGAGACCCCTTGAGCTGAGTGCTTTGCTCTGTCATACCTTTCTGGCCTCATATGGG 240
 DB 4180 CTGGAGACCCCTTGAGCTGAGTGCTTTGCTCTGTCATACCTTTCTGGCCTCATATGGG 4239
 QY 241 GCTTGGCCATTGTCCCTTCACCTCAGATCTCTCTTTTCAGTGCCAGGAAGTCATCTTGA 300
 DB 4240 GCTTGGCCATTGTCCCTTCACCTCAGATCTCTCTTTTCAGTGCCAGGAAGTCATCTTGA 4299
 QY 301 ACTTAAGCTTTTCAGAGCCCCCTTCAGTTTTCAGTCTCTAGAGGTGGACCTTCGATT 360
 DB 4300 ACTTAAGCTTTTCAGAGCCCCCTTCAGTTTTCAGTCTCTAGAGGTGGACCTTCGATT 4359
 QY 361 CCTTTGCTCTGTGCCCTGTAGCCCTCAGGTCCAGGCTTAAGCAAGGTCTCCTCACCTGGC 420
 DB 4360 CCTTTGCTCTGTGCCCTGTAGCCCTCAGGTCCAGGCTTAAGCAAGGTCTCCTCACCTGGC 4419
 QY 421 CTGGGAGAGTCCCAGGAGCTGCAGTGCCTGTGCGGGTAGGATGCTGATGCCAGATT 480
 DB 4420 CTGGGAGAGTCCCAGGAGCTGCAGTGCCTGTGCGGGTAGGATGCTGATGCCAGATT 4479
 QY 481 TCCCGTTAGAGAGCCCTTCCCTATCCTCAGCGGCTCTAGCTTTGTGTTACTTACTTTGTT 540
 DB 4480 TCCCGTTAGAGAGCCCTTCCCTATCCTCAGCGGCTCTAGCTTTGTGTTACTTACTTTGTT 4539
 QY 541 CCACCTTTAATCCAAATGTACCACGACACCAAGTGTGCACAGTCTCTCTGGGGTTTCAGG 600
 DB 4540 CCACCTTTAATCCAAATGTACCACGACACCAAGTGTGCACAGTCTCTCTGGGGTTTCAGG 4599
 QY 601 AGGGATGTAAGACATACCCCTTGCCCTTACGGCACTATGGCCAGAGGGGGCAGTGCAC 660
 DB 4600 AGGGATGTAAGACATACCCCTTGCCCTTACGGCACTATGGCCAGAGGGGGCAGTGCAC 4659
 QY 661 TAGGCAGAGCGCGGAGCCAGCAGATGGATACACTCAGAGGACCTTCAGCAGGCGAGAG 720
 DB 4660 TAGGCAGAGCGCGGAGCCAGCAGATGGATACACTCAGAGGACCTTCAGCAGGCGAGAG 4719
 QY 721 GCAGGGAGAGGAGGAGGTCTACAGTTCTGCG 751
 DB 4720 GCAGGGAGAGGAGGAGGTCTACAGTTCTGCG 751

QY 211 CTTGCATCTCTTCTGCTCCTAGTAGTGGGCTTGGCCATTGCTCCCTACTCCAGATCT 270
Db 1661 CTTGCATCTCTTCTGCTCCTAGTAGTGGGCTTGGCCATTGCTCCCTACTCCAGATCT 1720
QY 271 CTTCTTCAGGTCAGGAGTGCAT 295
Db 1721 CTTCTTCAGGTCAGGAGTGCAT 1745
RESULT 3
ID A3N24174 standard; cDNA; 256 BP.
XX AC ABN24174;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX polynucleotide sequence SEQ ID NO:16825.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimketa RA, Leach MD;
XX WPI; 2002-106308/14.
XX P-PSDB; ABP08422.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders
XX PS Disclosure; SEQ ID 16825; 1037pp; English.
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where SEQ ID NO:1-11491 (see Table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from

CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 256 BP; 33 A; 93 C; 79 G; 51 T; 0 other;
Query Match 2.9%; Score 22; DB 24; Length 256;
Best Local Similarity 100.0%; Pred. No. 0.69; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Mismatches 0; Gaps 0;
QY 697 CAGAGGAGCCTGCAGCAGGCAG 718
Db 233 CAGAGGAGCCTGCAGCAGGCAG 212
RESULT 4
ID ABL18235/C
XX ABL18235 standard; DNA; 903 BP.
XX AC ABL18235;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6178.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions
XX PS Claim 1; SEQ ID NO 6178; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 903 BP; 207 A; 266 C; 245 G; 185 T; 0 other;
Query Match 2.7%; Score 20; DB 23; Length 903;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 702 GAGCCTGCAGCAGGCAGG 721
Db 58 GAGCCTGCAGCAGGCAGG 39

```

RESULT 5
AAX31841/c
ID AAX31841 standard; DNA; 1347 BP.
XX AC AAX31841;
XX AC AAX31841;
DT 11-JUN-1999 (first entry)
XX DE Human T-Box polypeptide, Tbx19 encoding DNA.
XX KW T-Box polypeptide; Tbx19; Holt-Oram syndrome; Ulnar Mammmary syndrome;
KW DiGeorge syndrome; cancer; psychiatric disorder; heart defect; human; ss.
XX OS Homo sapiens.
XX PN WO9911783-A2.
XX PD 11-MAR-1999.
XX PF 02-SEP-1998; 98WO-EP05713.
XX PR 17-AUG-1998; 98GB-0017900.
XX PR 03-SEP-1997; 97EP-0306830.
XX PA (SMK) SMITHKLINE BEECHAM PLC.
XX PA (UYNO-) UNIV NOTTINGHAM.
XX PI Brook JD, Terrett JA, Yi CH;
XX DR WPI: 1999-205187/17.
XX DR P-PSDB; RAY03773.
XX PT Tbx19 protein - used to treat Holt-Oram syndrome, Ulnar Mammmary
XX PT syndrome, DiGeorge syndrome, cancer, psychiatric disorders, and
XX PT heart defects
XX PS Claim 9; Page 23; 32pp; English.
XX CC This DNA encodes a T-Box polypeptide having Tbx19 activity. Host cells
XX CC containing a vector comprising the Tbx19 nucleic acids are used for the
XX CC recombinant expression of the protein. The Tbx19 polypeptide can be used
XX CC for treating Holt-Oram syndrome, Ulnar Mammmary syndrome, DiGeorge
XX CC syndrome, cancer, psychiatric disorders, and heart defects.
XX SQ Sequence 1347 BP; 317 A; 404 C; 339 G; 287 T; 0 other;

Query Match 2.7%; Score 20; DB 20; Length 1347;
Best Local Similarity 100.0%; Pred. No. 6.9; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 710 AGCAGGCAGGCAGGCAGGAG 729
Db 816 AGCAGGCAGGCAGGCAGGAG 797
|||||

RESULT 6
AAT58853/c
ID AAT58853 standard; cDNA; 2728 BP.
XX AC AAT58853;
XX AC AAT58853;
DT 02-APR-1997 (first entry)
XX DE cDNA encoding high affinity Na+-dependent, L-Proline transporter.
XX DE High-affinity sodium dependent L-Proline transporter; rat brain;
KW rTN2-2-20; neurotransmitter transporter; ss.
XX OS Rattus rattus.
XX OS Rattus rattus.
XX FH Key Location/Qualifiers

Query Match 2.7%; Score 20; DB 18; Length 2728;
Best Local Similarity 100.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 708 GCAGCAGGCAGGCAGGCAGG 727
Db 757 GCAGCAGGCAGGCAGGCAGG 738
|||||

RESULT 7
AAV28861/c
ID AAV28861 standard; cDNA; 2728 BP.
XX AC AAV28861;
XX AC AAV28861;
DT 04-AUG-1998 (first entry)
XX DE Rat proline transporter rPROT encoding cDNA.
XX KW Rat; proline transporter; rPROT; neurotransmission; CNS; L-proline;
XX KW central nervous system; glutamatergic pathway; sodium; ss.
XX OS Rattus sp.
XX OS Rattus sp.
XX FH Key Location/Qualifiers
XX FT 1..2004
XX FT /*tag= a
XX FT /product= "proline transporter"
XX FT /transl_except= (pos:1..3,aa:Pro)
XX FT /transl_except= (pos:4..6,aa:Arg)
XX FT /transl_except= (pos:7..9,aa:Leu)
XX FT /transl_except= (pos:10..12,aa:ile)
XX FT /transl_except= (pos:13..15,aa:Asn)

```

FT /transl_except= (pos:16..18,aa:Glu)
 FT /transl_except= (pos:19..21,aa:Thr)
 FT /transl_except= (pos:22..24,aa:Arg)
 FT /transl_except= (pos:25..27,aa:Ala)
 FT /transl_except= (pos:28..30,aa:Asn)
 FT /transl_except= (pos:31..33,aa:Ser)
 FT /transl_except= (pos:34..36,aa:Pro)
 FT /transl_except= (pos:37..39,aa:Arg)
 FT /transl_except= (pos:40..42,aa:Thr)
 FT /transl_except= (pos:43..45,aa:Glu)
 FT /transl_except= (pos:46..48,aa:Arg)
 FT /transl_except= (pos:49..51,aa:Pro)
 FT /transl_except= (pos:52..54,aa:Glu)
 FT /transl_except= (pos:55..57,aa:Pro)
 FT /transl_except= (pos:58..60,aa:Thr)
 FT /transl_except= (pos:61..63,aa:Ile)
 FT /transl_except= (pos:64..66,aa:Asp)
 FT /transl_except= (pos:67..69,aa:Glu)
 FT /transl_except= (pos:70..72,aa:Ser)
 FT /transl_except= (pos:73..75,aa:Glu)
 FT /transl_except= (pos:76..78,aa:Gln)
 FT /transl_except= (pos:79..81,aa:Glu)
 FT /transl_except= (pos:82..84,aa:Asn)
 FT /transl_except= (pos:85..87,aa:Cys)
 FT /transl_except= (pos:88..90,aa:Glu)

XX US5759788-A.

PN 02-JUN-1998.

XX 03-DEC-1996; 96US-0753985.

XX 01-MAY-1992; 92US-0879617.

PR 03-DEC-1996; 96US-0753985.

XX (UYDU-) UNIV DUKE.

PA (UYEM-) UNIV EMORY.

XX Blakely RD, Caron MG, Fremerey RT;

XX WPI; 1998-332127/29.

DR P-PSDB; AAW57224.

XX L-proline transporter polypeptide - used for study of the
 PT transporter and antibody against it for detection of the transporter

XX Claim 1; Column 25-28; 37pp; English.

XX The present sequence encodes rat proline transporter (rPROT), which is
 CC a high-affinity, Na+-dependent L-proline transporter expressed in the
 CC brain. The present invention also describes an antibody that is
 CC specifically reactive with the L-proline transporter and does not
 CC cross-react with any other neurotransmitter transporter. The antibody
 CC can be used in assays for detecting the L-proline transporter in samples
 CC containing one or more proteins. The DNA encoding the polypeptide can
 CC be used for recombinant production of the transporter. The polypeptide
 CC can be used to make antibodies, in research studies, and in the design
 CC of modulating compounds for clinical treatments. This is useful as the
 CC transporter is thought to be involved in the release of amino acid
 CC transmitters for excitatory pathways in the central nervous system. The
 CC products can be used to study function and disorders involving these
 CC transporters.

XX Sequence 2728 BP; 555 A; 787 C; 753 G; 633 T; 0 other;

Query Match 2.7%; Score 20; DB 19; Length 2728;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 GCAGCAGCAGCAGCAGG 727

|||||

Db 757 GCAGCAGCAGCAGCAGG 738

RESULT 8

ABLI8234

ID ABLI8234 standard; DNA; 3129 BP.

XX ABLI8234;

AC ABLI8234;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6175.

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6175.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 6175; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI16175) and the encoded proteins
 CC (ABB57737-ABB572072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3129 BP; 866 A; 636 C; 657 G; 970 T; 0 other;

Query Match 2.7%; Score 20; DB 23; Length 3129;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGCCTGCAGCAGCAGG 721

|||||

Db 2072 GAGCCTGCAGCAGCAGG 2091

RESULT 9

ABN58123/c

ID ABN58123 standard; DNA; 65 BP.

XX ABN58123;

XX 15-JUL-2002 (first entry)

XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:30871.

XX Human; mouse; rat; splice transcript; detection; RNA transcript;

XX splice variant; transcriptome; oligonucleotide library; ss.

OS Mus musculus.
 XX WO200210449-A2.
 PN
 XX
 PD 07-FEB-2002.
 XX
 XX 20-JUL-2001; 2001WO-1501903.
 XX
 XX 28-JUL-2000; 2000US-221607P.
 PR
 XX 02-MAY-2001; 2001US-287724P.
 XX
 XX (COMP-) COMPUTEN INC.
 PA
 XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX WPI; 2002-257383/30.
 XX
 XX New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes.
 XX
 XX Example 1; SEQ ID 30871; 47pp; English.
 XX
 XX The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterizing the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 65 BP; 19 A; 17 C; 17 G; 12 T; 0 other;
 Query Match 2.5%; Score 19; DB 24; Length 65;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 296 CTTGAACCTTAACCTTCCAG 314
 Db 48 CTTGAACCTTAACCTTCCAG 30
 RESULT 10
 AAA00717/c
 ID AAA00717 standard; cDNA; 298 BP.
 XX
 AC AAA00717;
 XX
 XX 19-MAY-2000 (first entry)
 DT
 XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:708.
 DE
 XX Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;

KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX Homo sapiens.
 OS
 XX WO9958675-A2.
 PN
 XX 18-NOV-1999.
 PD
 XX 13-MAY-1999; 99WO-US10602.
 PF
 XX 14-MAY-1998; 98US-0085426.
 PR
 XX 15-MAY-1998; 98US-0085537.
 PR
 XX 15-MAY-1998; 98US-0085696.
 PR
 XX 21-OCT-1998; 98US-0105234.
 PR
 XX 27-OCT-1998; 98US-0105877.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX (HYSE-) HYSEQ INC.
 PA
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX WPI; 2000-126369/11.
 DR
 XX Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PT
 XX Claim 1; Page 353; 1097pp; English.
 PS
 XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SQ Sequence 298 BP; 41 A; 64 C; 77 G; 91 T; 25 other;
 Query Match 2.4%; Score 18; DB 21; Length 298;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 AGGACTGCCAACCCCTC 37
 Db 57 AGGACTGCCAACCCCTC 40
 RESULT 11
 AAT70809
 ID AAT70809 standard; DNA; 438 BP.
 XX
 AC AAT70809;
 XX
 XX 30-JUL-1997 (first entry)
 DT
 XX Mouse anti-idiotypic antibody 3B6 light chain variable region.
 DE
 XX Anti-idiotypic; anti-EGFR; epidermal growth factor receptor; tumour;
 KW cancer; neoplasia; glioma; melanoma; carcinoma; drug manufacture; ss.
 KW


```

OS Mus musculus.
XX key Location/Qualifiers
XX CDS 1..438
XX /tag= a
XX /note= "no stop codon given"
XX sig_peptide 1..60
XX /tag= b
XX primer_bind 1..29
XX /tag= c
XX primer_bind 416..438
XX /tag= d
XX EP745612-A1.
XX
XX PD 04-DEC-1996.
XX
XX PF 14-MAY-1996; 96EP-0107651.
XX
XX PR 26-MAY-1995; 95EP-0107967.
XX
XX PA (MERE ) MERCK PATENT GMBH.
XX
XX PI Adan J, Carceller A, Gomez A, Piulats J, Rosell E;
XX
XX DR WPI: 1997-013659/02.
XX
XX DR P-PSDB; AAM19578.
XX
XX PT Monoclonal anti-idiotypic antibodies mimicking epidermal growth
XX factor receptor - useful for tumour therapy
XX
XX PS Claim 6; Fig 5D; 28pp; English.
XX
XX CC Murine monoclonal anti-idiotypic antibodies, 15B8, 3B6 and 5A6 are
XX new. They induce an immune response against epidermal growth factor
XX receptor (EGFR). The sequences of the heavy and light chain variable
XX regions of these antibodies are given in the specification. The
XX antibodies are used for the manufacture of drugs directed against
XX tumours that express EGFR on their surface, including melanomas,
XX gliomas and carcinomas.
XX
XX SQ Sequence 438 BP; 105 A; 123 C; 109 G; 101 T; 0 other;

Query Match 2.4%; Score 18; DB 18; Length 438;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AGGGAGGCTACACGTTTC 748
| | | | | | | | | | | | | | | | | |
Db 346 AGGGAGGCTACACGTTTC 363

RESULT 12
AA572980/c
ID AA572980 standard; cDNA; 588 BP.
XX
XX AC AA572980;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #8784.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR

31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Dmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
P-PSDB; ABG08793.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity
Claim 1; SEQ ID No 8784; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pcr_sequences.
XX
XX SQ Sequence 588 BP; 160 A; 139 C; 145 G; 144 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 588;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 TCCTTGCATACCTCTTCT 226
| | | | | | | | | | | | | | | | | |
Db 460 TCCTTGCATACCTCTTCT 443

RESULT 13
AAC43758/c
ID AAC43758 standard; DNA; 1326 BP.
XX
XX AC AAC43758;
XX
XX DT 18-OCT-2000 (first entry)
XX
XX DE Zea mays DNA fragment SEQ ID NO: 40403.
XX
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic;
XX pathway; promoter; termination sequence; corn; ss.
XX
XX OS Zea mays subsp. mays.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.

```

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131440.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0143055.
PR 06-JUL-1999; 99US-0143290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158599.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.4%; Score 18; DB 21; Length 1326;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 AGGCAGAGGCAGGAGGAGA 730
DB 137 AGGCAGAGGCAGGAGGAGA 120
|||||

RESULT 14
AAS75162
ID AAS75162 standard; cDNA; 1427 BP.
XX
AC AAS75162;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10966.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT.
XX
XX WPI; 2001-639362/73.
XX
P-PSDB; ABG10975.
DR

```

```

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 10966; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1427 BP; 329 A; 365 C; 400 G; 333 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 1427;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 AGGCAGAGGCAGGAGGAGA 730
DB 1031 AGGCAGAGGCAGGAGGAGA 1048
|||||

RESULT 15
AAC59691
ID AAC59691 standard; cDNA; 1572 BP.
XX
AC AAC59691;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #13.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200056880-A1.
XX
PD 28-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US06781.
XX
PR 19-MAR-1999; 99US-0125363.
PR 08-DEC-1999; 99US-0169617.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-602220/57.
DR

```

DR P-PSDB; AAB34466.

XX Nucleic acid molecules encoding human secreted proteins, used in

PT preventing, treating or ameliorating disorders such as Parkinson's and

PT Alzheimer's diseases, cancers and infections.

XX

PS Claim 1; Pages 347-348; 422pp; English.

XX

CC The invention relate to the isolation of genes AAC59679-C59728 encoding

CC XX human secreted proteins AAB34454-B34503. The genes can be used to

CC generate fusion proteins by linking to the gene for the human

CC immunoglobulin G FC portion (AAC59670) for increasing the stability of

CC the fusion protein as compared to the human protein only. The genes and

CC proteins are useful for preventing, ameliorating or treating medical

CC conditions, e.g. by protein or gene therapy. The genes are isolated

CC from a range of human tissues disclosed in the specification. The

CC nucleic acids, proteins, antibodies and (ant)agonists are useful in

CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

CC and ovarian cancer, and other cancers of the adrenal gland, bone,

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune

CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative

CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)

CC wound healing; (e) neurological diseases e.g. cerebral anoxia and

CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.

XX

XX Sequence 1572 BP; 333 A; 378 C; 440 G; 421 T; 0 other;

XX

Query Match 2.4%; Score 18; DB 21; Length 1572;

Best Local Similarity 100.0%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 AGAGGCAGGAGGAGGG 734

|||||

Db 951 AGAGGCAGGAGGAGGG 968

Search completed: April 12, 2003, 03:53:55

Job time : 136.291 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 02:50:35 ; Search time 824.43 Seconds
(without alignments)
14752.993 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750

Perfect score: 751

Sequence: 1 aaagcagacttagcactca.....gggagggtotacacgttctgc 751

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154065 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pin:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	230	30.6	523	17	AZ519305
C 2	21	2.8	483	13	BM482706
C 3	20	2.7	124	9	AA873760
C 4	20	2.7	165	17	A2731372
C 5	20	2.7	208	12	BF176940
C 6	20	2.7	282	13	BI141046

7	20	2.7	322	10	BB464276
8	20	2.7	345	10	AW347859
9	20	2.7	361	10	BB873702
C 10	20	2.7	389	10	BB872026
C 11	20	2.7	436	9	AA984230
C 12	20	2.7	447	12	BF177173
C 13	20	2.7	448	12	BF109365
C 14	20	2.7	550	14	BQ208223
C 15	20	2.7	558	12	BG557892
C 16	20	2.7	591	13	BM487057
C 17	20	2.7	654	13	BI067137
C 18	20	2.7	663	13	BM427480
C 19	20	2.7	688	17	AZ945331
C 20	20	2.7	737	9	AL584530
C 21	20	2.7	966	10	AW982491
C 22	20	2.7	1012	17	CNSQ3NPW
C 23	20	2.7	1101	17	CNSQ00M2U
C 24	20	2.7	1747	12	BF315159
C 25	19	2.5	152	9	AI910314
C 26	19	2.5	206	13	BM030031
C 27	19	2.5	284	14	R42226
C 28	19	2.5	296	10	AW345495
C 29	19	2.5	298	17	BH052505
C 30	19	2.5	310	10	BE477522
C 31	19	2.5	311	12	BE984851
C 32	19	2.5	334	14	F10476
C 33	19	2.5	337	10	AW512241
C 34	19	2.5	344	17	AZ717051
C 35	19	2.5	351	10	BE477529
C 36	19	2.5	390	14	R44562
C 37	19	2.5	420	17	BH029153
C 38	19	2.5	497	14	R42678
C 39	19	2.5	531	12	BG649562
C 40	19	2.5	575	17	B53005
C 41	19	2.5	582	13	BM439558
C 42	19	2.5	585	17	AQ464274
C 43	19	2.5	610	9	AL676756
C 44	19	2.5	615	9	AL657815
C 45	19	2.5	673	12	BG854024

ALIGNMENTS

RESULT 1

AZ519305/C

LOCUS

DEFINITION

DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ519305 523 bp DNA linear GSS 16-OCT-2000
RPCI-11-91H7.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-91H7,
DNA sequence.

AZ519305

AZ519305.1 GI:10830677

GSS.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 523)

Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

J.C.

BAC end sequences of library RPCI-11

Unpublished (1997)

Other_GSSs: RPCI11-91H7.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

BB464276 BB464276
AW347859 32150 MAR
BB873702 BB873702
BB872026 BB872026
AA984230 am82903.s
BF177173 BM1_2_F05
BF109365 715h11.x
BQ208223 UI-R-EPO-
BG557892 EM1_55_H0
BM487057 pgm2n.pk0
BI067137 pglfn.pk0
BM427480 pglfn.pk0
A2945391 2M0206E14
AL584530 AL584530
AW982491 HVSMEG000
AL252365 Tetraodon
AL079066 Drosophila
BF315159 601899519
AI910314 QV-BT247-
BM030031 488526 MA
R42226 yf90d11.s1
AW345495 15006 MAR
BH052505 RPCI-24-3
BE477522 161263 BA
BE984851 UI-M-CGOp
F10476 HSC3FF082.n
AW512241 xU55802.x
AZ717051 RPCI-24-1
BE477529 161275 BA
R44562 y929f08.s1
BH029153 RPCI-24-3
R42678 y913b08.s1
BG649562 EM1_80_H0
B53005 CIT-HSP-200
BM439558 pgr1c.pk0
AQ464274 HS_3101.A
AL676756 AL676756
AL657815 AL657815
BG854024 1024038G0

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904 e. Vector identified by cross_match with the -minscore 18 and -micmatch 12 options.
PCR Primers

```

BASE COUNT      27 a      37 c      30 g      30 t
ORIGIN
M. Fatima Bonaldo. "
Query Match      2.7%; Score 20; DB 9; Length 124;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 ACCCTCCCGGTTTCCTTG 51
|||||
Db 76 ACCCTCCCGGTTTCCTTG 95

RESULT 4
AZ731372
LOCUS
DEFINITION      RP1-24-128A17.TJ RPCI-24 Mus musculus genomic clone RPCI-24-128A17
                  , DNA sequence.
ACCESSION      AZ731372
VERSION
KEYWORDS      GSS
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 165)
Tsedaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Zhao,S., Nierman,M., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsedaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Other_GSSs: RPCI-24-128A17.TV
Contact: Shaving Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
Page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 128 row: A column: 17
Seq primer: SP6
Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..165
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-128A17"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      50 a      31 c      50 g      34 t
ORIGIN

Query Match      2.7%; Score 20; DB 17; Length 165;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 GCAGGCAGAGCAGAGGAGA 730
|||||
Db 132 GCAGGCAGAGCAGAGGAGA 151

RESULT 5
BF176940/c
LOCUS
DEFINITION      EM1.4.F05.bl_A002 Embryo 1 (EMI) Sorghum bicolor cDNA, mRNA
                  208 bp mRNA linear EST 31-OCT-2000
                  sequence.
ACCESSION      BF176940
VERSION
KEYWORDS      EST.
SOURCE      sorghum.
ORGANISM      Sorghum bicolor
REFERENCE
AUTHORS      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 208)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence start: 5
High quality sequence stop: 194
POLYA-No.
Location/Qualifiers
1..208
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EMI)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT      24 a      88 c      59 g      37 t
ORIGIN

Query Match      2.7%; Score 20; DB 12; Length 208;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 GCAGAGCAGAGGAGGCG 734
|||||
Db 39 GCAGAGCAGAGGAGGCG 20

RESULT 6
BF141046/c
LOCUS
DEFINITION      IP1.42.E01.bl_A002 Immature pannicle 1 (IPI) Sorghum bicolor cDNA,
                  mRNA sequence.
ACCESSION      BF141046
VERSION
KEYWORDS      EST.
SOURCE      sorghum.
ORGANISM      Sorghum bicolor
REFERENCE
AUTHORS      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 282)
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt
L.H.
An EST database from Sorghum: developing preanthesis pannicles
Unpublished (2001)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

```


USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smithemail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 9 row: L column: 8

Seq primer: ATTTAGGTGACACTAG.

Location/Qualifiers

1..345

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC lBOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;

Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

88 a 68 c 103 g 86 t

BASE COUNT

ORIGIN

Query Match 2.7% Score 20; DB 10; Length 345;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 585 TCTCTGGGTTTCAGGAGG 604

|||||

Db 104 TCTCTGGGTTTCAGGAGG 123

RESULT 9

LOCUS

BB873702 RIKEN full-length enriched, pooled tissues, intestinal

mucosa, etc. Mus musculus cDNA clone G630049K20 5', mRNA sequence.

BB873702

BB873702.1 GI:17119912

EST.

KEYWORDS

SOURCE

ORGANISM

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 361)

Akimura.T., Arakawa.T., Carninci.P., Furuno.M., Hanagaki.T.,

Hayatsu.N., Hiramoto.K., Hiraoka.T., Hirozane.T., Imotani.K., Ishii

,Y., Ito.M., Kawai.J., Kojima.Y., Konno.H., Kouda.M., Matsuyama.T.,

Nakamura.M., Nishi.K., Nomura.K., Numasaki.R., Okazaki.Y., Okido.T.,

Saito.R., Sakai.C., Sakai.K., Sakazume.N., Sasaki.D., Sato.K.,

Shibata.K., Shinagawa.A., Shiraki.T., Sogabe.Y., Suzuki.H., Tagawa

,A., Takahashi.F., Takaku-Akahira.S., Tanaka.T., Tomaru.A., Toya.T.

, Watahiki.A., Yasunishi.A., Muramatsu.M. and Hayashizaki.Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura.T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci.P., Shibata.Y., Hayatsu.N., Sugahara.Y., Shibata.K., Itoh

.M., Konno.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

FEATURES

source

Location/Qualifiers

1..361

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone_lib="G630049K20"

/clone_lib="RIKEN full-length enriched, pooled tissues,

intestinal mucosa, etc."

/note="pooled tissues; (tissue_type=intestinal mucosa,

dev_stage=adult, sex=male); (tissue_type=accessory

axillary lymph node, dev_stage=adult, sex=male);

(tissue_type=jejunal and colic lymph node, dev_stage=adult,

sex=male); (tissue_type=vesicular gland, dev_stage=adult,

sex=male); (tissue_type=spinal cord, dev_stage=11 days

embryo); (tissue_type=brain, dev_stage=13 days embryo),

(tissue_type=spinal cord, dev_stage=15 days embryo),

(tissue_type=lung, dev_stage=15 days embryo),

(tissue_type=brain, dev_stage=15 days embryo),

(tissue_type=ovary and uterus, dev_stage=10 days pregnant

adult, sex=female); (tissue_type=cortex, dev_stage=0 day

neonate); (tissue_type=cerebellum, dev_stage=1 month

neonate); (tissue_type=diencephalon, dev_stage=16 days

neonate, sex=male); (tissue_type=medulla oblongata,

dev_stage=16 days neonate, sex=male);

(tissue_type=cerebellum, dev_stage=21 days neonate);

(tissue_type=testis, dev_stage=8 days neonate, sex=male)"

91 a 102 c 111 g 57 t

BASE COUNT

ORIGIN

Query Match 2.7% Score 20; DB 10; Length 361;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 GCTCACCAGCTCTCTCCA 164

|||||

Db 319 GCTCACCAGCTCTCTCCA 300

RESULT 10

LOCUS

BB872026

DEFINITION

musculus

ACCESSION

BB872026

VERSION

BB872026.1

KEYWORDS

EST.

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 389)

Akimura.T., Arakawa.T., Carninci.P., Furuno.M., Hanagaki.T.,

Hayatsu.N., Hiramoto.K., Hiraoka.T., Hirozane.T., Imotani.K., Ishii

,Y., Ito.M., Kawai.J., Kojima.Y., Konno.H., Kouda.M., Matsuyama.T.,

Nakamura.M., Nishi.K., Nomura.K., Numasaki.R., Okazaki.Y., Okido.T.,

Saito.R., Sakai.C., Sakai.K., Sakazume.N., Sasaki.D., Sato.K.,

Shibata.K., Shinagawa.A., Shiraki.T., Sogabe.Y., Suzuki.H., Tagawa

wagi.K., Fujiwaki.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E.,
Watahiki.M., Yoneda.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsuyama
.S., Kawai.J., Okazaki.Y., Muramatsu.M., Inoue.Y., Kira.A. and
Hayashizaki.Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno.H., Fukunishi.Y., Shibata.K., Itoh.M., Carninci.P., Sugahara

.Y. and Hayashizaki.Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akinura, T., et al. 2001)

TITLE
JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source

Location/Qualifiers
1. .389
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630034B09"
/clone_lib="RIKEN full-length enriched, adult male gall bladder"
/sex="male"
/tissue_type="gall bladder"
/dev_stage="adult"
/note="pooled tissues : (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed),"
95 a 115 c 117 g 62 t

BASE COUNT
ORIGIN

Query Match 2.7%; Score 20; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 GCTCACACGCTCTCTCCA 164
Db 320 GCTCACACGCTCTCTCCA 301
|||||

RESULT 11
AA984230
LOCUS

DEFINITION
am82903.sl Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:1629652 3' similar to gb|M32845|SEU28SRRNA Spea multiplicate 28S ribosomal RNA, (tRNA); gb:M65105 SODIUM-DEPENDENT NORADRENALINE TRANSPORTER (HUMAN); mRNA sequence.
AA984230

ACCESSION
VERSION

AA984230.1 GI:3162755

KEYWORDS
SOURCE
ORGANISM

EST.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 436)

REFERENCE

AUTHORS

Kriznan, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)

TITLE

JOURNAL

COMMENT

Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. RT from Amersham

High quality sequence stop: 435.

FEATURES

source

Location/Qualifiers
1. .436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1629652"
/clone_lib="Stratagene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site 1: EcoRI; Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Volken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of Individuals with Psychiatric Diseases (Unpublished) Stanley Neurovirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD."
87 a 120 c 141 g 85 t

BASE COUNT
ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 708 GCACGAGGACGAGGAGG 727
Db 156 GCACGAGGACGAGGAGG 175
|||||

RESULT 12
BF177173/c

LOCUS

DEFINITION
EM1_2_F05.bl_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
447 bp mRNA linear EST 31-OCT-2000
sequence.

ACCESSION

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 593 0210
Email: mmpatr@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 427
POLYA-No.

FEATURES

source

Location/Qualifiers

1. 447

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Embryo.1 (EM1)"

/note="Organ: Embryos germinated for 24 hr; Vector: pBlueScript II from Lambda zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 86 a 174 c 114 g 73 t

ORIGIN

Query Match 2.7%; Score 20; DB 12; Length 447;

Best Local Similarity 100.0%; Pred. No. 71;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 GCAGGCGAGGAGGAGG 734

11111111111111111111

Db 98 GCAGGCGAGGAGGAGG 79

RESULT 13

BF109365/c

LOCUS

DEFINITION 715h11.x1 Soares.NSF.F8.9W_OT_PA_P.S1 Homo sapiens CDNA clone

IMAGE:3524997 3' similar to TR:Q9JZ7 Q9JZ7 E2RIN ; contains

MER2 t2 MER2 repetitive element ; mRNA sequence.

ACCESSION BF109365

VERSION BF109365.1 GI:10939055

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -400P from Gibco

High quality sequence stop: 276.

Location/Qualifiers

1. 448

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3524997"

/clone_lib="Soares.NSF.F8.9W_OT_PA_P.S1"

/lab_host="CH10B"

/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro.

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was from

PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following

libraries and clones: Soares NbHSF pool 1;

309384-310919, 323208-325895 Soares Nb2HP pool 1;
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1;
758280-760583, 772104-774407 Soares NbHPA pool 1;
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1; 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo.*
BASE COUNT 103 a 83 c 110 g 152 t

ORIGIN

Query Match

Best Local Similarity

Matches 20; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 146 CTCACGAGCTTCTCTCCAC 165

11111111111111111111

Db 263 CTCACGAGCTTCTCTCCAC 244

11111111111111111111

RESULT 14

BQ208223

LOCUS

DEFINITION

BQ208223

UI-R-EP0-coa-j-10-0-UI.s1 UI-R-EP0 Rattus norvegicus CDNA clone

EST.

ACCESSION BQ208223.1 GI:20424693

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dt track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. the sequence tag served to identify it as a clone from the

normalized duodenum library cDNA Library Preparation: M.B. Soares

Lab Clone distribution: clones will be available through Research

Genetics (www.resgen.com) The following repetitive elements were

found in this cDNA sequence: 14-85, >MSTC#LTR/MaLR

Seq primer: M13 Forward

POLYA=yes.

Location/Qualifiers

1. 550

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-EP0-coa-j-10-0-UI"

/clone_lib="UI-R-EP0"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-EP0 is a

subtracted cDNA library constructed according to Bonaldo,

Lennon and Soares, Genome Research, 6:791-806, 1996. First

strand cDNA synthesis was primed with an oligo-dt primer

containing a Not I site. Double stranded cDNA was ligated

to an EcoR I adaptor, digested with Not I, and cloned

directionally into p7T3-Pac vector. The oligonucleotide

used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (df)18 tail. The sequence tags for this library are: distal colon, GAAGTGCTCC; osteoblast, AAGATATCAA; cell line R3327 5A, GGACTAGATC; cell line R3327 5P, CACGTGAGAT; duodenum, TGTGTTTCAT; prostate, CCAGG.

TAG_LIB-UI-R-EPO
TAG_TISSUE=duodenum
TAG_SEQ=GTGTTTCAT*

BASE COUNT 144 a 128 c 136 g 142 t

Query Match 2.7%; Score 20; DB 14; Length 550;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 706 CTGCAGCGCAGCAGCAGA 725
|||||
Db 380 CTGCAGCGCAGCAGCAGA 399

RESULT 15
BG557892/c
LOCUS EML_55_H08.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA 558 bp linear EST 10-APR-2001
DEFINITION

ACCESSION BG557892
VERSION BG557892.1 GI:13586890
KEYWORDS EST.

SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 558)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.

REFERENCE An EST database from Sorghum: developing embryos
AUTHORS Unpublished (2000)
TITLE Contact: Cordonnier-Pratt MM
JOURNAL Laboratory for Genomics and Bioinformatics
COMMENT The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 533
POLYA=No.

FEATURES

Location/Qualifiers
1..558
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site 1: xhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 111 a 207 c 138 g 102 t

Query Match 2.7%; Score 20; DB 12; Length 558;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 715 GCAGAGCGCAGCAGCAGGG 734
|||||
Db 98 GCAGAGCGCAGCAGCAGGG 79

Search completed: April 12, 2003, 08:59:13
Job time : 830.43 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 03:59:25 ; Search time 20.8447 Seconds
(without alignments)
11049.076 Million cell updates/sec

Title: US-09-914-152-3_COPY_4000_4750
Perfect score: 751
Sequence: 1 aaagcagactctagcactca.....gggagggtctacagttctgc 751

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	2.7	2728	1	US-07-879-617A-7
C 2	20	2.7	2728	1	US-08-753-985-7
C 3	18	2.4	438	2	US-08-653-402B-11
C 4	18	2.4	6386	2	US-08-483-376-1
C 5	17	2.3	514	2	US-08-966-316-14
C 6	17	2.3	3357	3	US-08-726-214-7
C 7	17	2.3	4421	2	US-08-257-963B-9
C 8	17	2.3	4421	4	US-08-367-841A-9
C 9	17	2.3	4421	4	US-08-530-373D-6
C 10	17	2.3	4421	5	PCT-US95-07201-9
C 11	16	2.1	419	1	US-08-486-013-59
C 12	16	2.1	419	2	US-08-482-279-59
C 13	16	2.1	419	2	US-08-342-268-59
C 14	16	2.1	419	3	US-09-015-968-59
C 15	16	2.1	419	3	US-09-397-386-59
C 16	16	2.1	432	1	US-08-423-383-19
C 17	16	2.1	432	2	US-08-437-353A-19
C 18	16	2.1	490	4	US-08-896-164-66
C 19	16	2.1	575	4	US-09-385-982-123
C 20	16	2.1	675	2	US-08-114-555A-9
C 21	16	2.1	675	2	US-08-114-555A-10
C 22	16	2.1	675	3	US-08-559-397A-15
C 23	16	2.1	720	3	US-08-114-555A-11
C 24	16	2.1	720	3	US-08-559-397A-16
C 25	16	2.1	738	4	US-09-392-184-13
C 26	16	2.1	784	4	US-08-896-164-56
C 27	16	2.1	810	4	US-08-406-030A-18

28 16 2.1 1265 1 US-07-816-283-3 Sequence 3, Appli
29 16 2.1 1265 1 US-08-417-103-3 Sequence 3, Appli
30 16 2.1 1334 2 US-08-916-901-2 Sequence 2, Appli
31 16 2.1 1334 4 US-09-154-602-2 Sequence 2, Appli
32 16 2.1 1490 4 US-08-964-127-3 Sequence 3, Appli
33 16 2.1 1490 4 US-09-496-692-3 Sequence 3, Appli
34 16 2.1 1607 4 US-09-374-454-5 Sequence 5, Appli
35 16 2.1 1730 6 5223391-8 Patent No. 5223391
36 16 2.1 1959 4 US-08-743-168B-39 Sequence 39, Appli
37 16 2.1 1959 4 US-08-743-168B-41 Sequence 41, Appli
38 16 2.1 1995 3 US-08-904-452-1 Sequence 1, Appli
39 16 2.1 1995 4 US-09-517-639-1 Sequence 1, Appli
40 16 2.1 2180 1 US-07-918-314-3 Sequence 3, Appli
41 16 2.1 2223 1 US-08-317-542A-8 Sequence 8, Appli
42 16 2.1 2223 1 US-08-439-818A-8 Sequence 8, Appli
43 16 2.1 2223 2 US-08-751-965-8 Sequence 8, Appli
44 16 2.1 2223 2 US-08-738-975-8 Sequence 8, Appli
45 16 2.1 2223 2 US-08-728-636-8 Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-07-879-617A-7/c
; Sequence 7, Application US/07879617A
; Patent No. 5580775
; GENERAL INFORMATION:
; APPLICANT: Freneau Jr., Robert T.
; APPLICANT: Caron, Marc G.
; APPLICANT: Blakely, Randy D.
; TITLE OF INVENTION: A High Affinity L-Proline Transporter
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/879,617A
; FILING DATE: 19920501
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2728 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rattus
; TISSUE TYPE: Brain
; IMMEDIATE SOURCE:
; LIBRARY: rat forebrain cDNA library
; CLONE: rTB2-2-20

US-07-879-617A-7

```
Query Match      2.7%; Score 20; DB 1; Length 2728;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 GCAGCAGCAGCAGCAGG 727
      |||||
DB 757 GCAGCAGCAGCAGCAGG 738

RESULT 2
US-08-753-985-7/c
; Sequence 7, Application US/08753985
; Patent No. 5759788
; GENERAL INFORMATION:
; APPLICANT: Frenau Jr., Robert T.
; APPLICANT: Caron, Marc G.
; APPLICANT: Blakely, Randy D.
; TITLE OF INVENTION: A High Affinity L-Proline Transporter
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,985
; FILING DATE: 03-DEC-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879617
; FILING DATE: 01-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Rattus
; TISSUE TYPE: Brain
; IMMEDIATE SOURCE:
; LIBRARY: rat forebrain cDNA library
; CLONE: rTB2-2-20
US-08-753-985-7

Query Match      2.7%; Score 20; DB 1;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 GCAGCAGCAGCAGCAGG 727
      |||||
DB 757 GCAGCAGCAGCAGCAGG 738

RESULT 3
US-08-653-402B-11

; Sequence 11, Application US/08653402B
; Patent No. 5969107
; GENERAL INFORMATION:
; APPLICANT: CARCELLER, Ana
; APPLICANT: ROSELL, Elisabeth
; APPLICANT: GOMEZ, Alicia
; APPLICANT: ADEN, Jaume
; APPLICANT: PIULATS, Jaume
; TITLE OF INVENTION: Anti-idiotypic antibodies which induce an
; TITLE OF INVENTION: Immune response against epidermal growth factor receptor.
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Milten, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Boulevard, Suite 1400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (SPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/653,402B
; FILING DATE: 24-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95107967.2
; FILING DATE: 26-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: MERCK 1781
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: monoclonal anti-idiotypic anti-EGFR antibody
; INDIVIDUAL ISOLATE: 3B6 light chain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..438
US-08-653-402B-11

Query Match      2.4%; Score 18; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AGGGAGGTCTACAGTTC 748
      |||||
DB 346 AGGGAGGTCTACAGTTC 363

RESULT 4
US-08-483-376-1/c
; Sequence 1, Application US/08483376
; Patent No. 5955330
; GENERAL INFORMATION:
; APPLICANT: Vasil, Vmila
; APPLICANT: Clancy, Maureen A.
; APPLICANT: Ferl, Robert J.
; APPLICANT: Vasil, Indra K.
```

APPLICANT: Haunah, L. C.
TITLE OF INVENTION: No. 5955330el Means for Enhancing Gene
TITLE OF INVENTION: Expression
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/483,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/418,540
FILING DATE: 07-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,115
FILING DATE: 04-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/830,956
FILING DATE: 05-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/353,854
FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 10-94B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays
STRAIN: Black Sweet
FEATURE:
NAME/KEY: exon
LOCATION: 131..182
FEATURE:
NAME/KEY: exon
LOCATION: 1211..1324
FEATURE:
NAME/KEY: exon
LOCATION: 1828..1948
FEATURE:
NAME/KEY: exon
LOCATION: 2041..2187
FEATURE:
NAME/KEY: exon
LOCATION: 2269..2460
FEATURE:
NAME/KEY: exon
LOCATION: 2605..2728
FEATURE:
NAME/KEY: exon
LOCATION: 2822..3038
FEATURE:

NAME/KEY: exon
LOCATION: 3256..3351
FEATURE:
NAME/KEY: exon
LOCATION: 3447..3620
FEATURE:
NAME/KEY: exon
LOCATION: 3702..3818
FEATURE:
NAME/KEY: exon
LOCATION: 3912..4078
FEATURE:
NAME/KEY: exon
LOCATION: 4158..4381
FEATURE:
NAME/KEY: exon
LOCATION: 4517..4835
FEATURE:
NAME/KEY: exon
LOCATION: 4768..5212
FEATURE:
NAME/KEY: exon
LOCATION: 5372..5510
FEATURE:
NAME/KEY: exon
LOCATION: 5636..5917
US-08-483-376-1
Query Match 2.4%; Score 18; DB 2; Length 6386;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 710 ACAGGCAGAGCAGGAGG 727
|||||
Db 5553 ACAGGCAGAGCAGGAGG 5536
RESULT 5
US-08-966-316-14/C
Sequence 14, Application US/08966316
Patent No. 5932445
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Au-Young, Janice
APPLICANT: Reddy, Roopa
APPLICANT: Murry, Lynn E.
APPLICANT: Mathur, Preeti
TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/966,316
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0424 US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 514 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TMLR3DT02
CLONE: 506333
US-08-966-316-14

Query Match 2.3%; Score 17; DB 2; Length 514;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 CTGCCAACACCTCCCC 40
|||||
Db 185 CTGCCAACACCTCCCC 169

RESULT 6

US-08-726-214-7
Sequence 7, Application US/08726214
Patent No. 6107076

GENERAL INFORMATION:

APPLICANT: Tang, Wei-Jen
APPLICANT: Gilman, Alfred G.
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD-450

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 3357 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-726-214-7

Query Match 2.3%; Score 17; DB 3; Length 3357;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 219 CTCCTTCTGGCCTCATA 235
|||||

Db 381 CTCCTTCTGGCCTCATA 397

RESULT 7

US-08-257-963B-9/c
Sequence 9, Application US/08257963B
Patent No. 5840686

GENERAL INFORMATION:

APPLICANT: Chader, Gerald J.; Becerra, S.
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki

TITLE OF INVENTION: PIGMENT EPITHELIUM

TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
TITLE OF INVENTION: AND EXPRESSING THE PROTEIN

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan & Finnegan

STREET: 345 Park Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/257,963B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/952,796

FILING DATE: 24-SEPT-1992

ATTORNEY/AGENT INFORMATION:

NAME: DOROTHY R. AUTH

REGISTRATION NUMBER: 36434

REFERENCE/DOCKET NUMBER: 20264126US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 4421 Base Pairs

TYPE: Nucleic Acid

STRANDEDNESS: Double

TOPOLOGY: Unknown

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Human

FEATURE:

NAME/KEY: JT101

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION: 7.1 kb Bam HI

OTHER INFORMATION: fragment Derived from human placental

OTHER INFORMATION: genomic DNA

US-08-257-963B-9

Query Match

Best Local Similarity 100.0%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 CAGTTTCCAGTCCTTA 341
|||||

Db 2508 CAGTTTCCAGTCCTTA 2492

RESULT 8

US-08-367-841A-9/c
Sequence 9, Application US/08367841A
Patent No. 6319687

RESULT 9
US-08-520-373D-6/c
; Sequence 6, Application US/08520373D
; Patent No. 6451763
; GENERAL INFORMATION:
; APPLICANT: Tombran-Tink, Joyce
; APPLICANT: Steele, Fintan R
; APPLICANT: Chader, Gerald J
; APPLICANT: Becerra, Sofia P

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20266412660

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4421 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: JTI
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.1 kb Bam HI
OTHER INFORMATION: fragment derived from human placental
OTHER INFORMATION: genomic DNA; Also referred to as JTI01
PCT-US93-07201-9

Query Match 2.3% Score 17; DB 5; Length 4421;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 CAGTTTCCAGTCCTTA 341
|||||
Db 2508 CAGTTTCCAGTCCTTA 2492

RESULT 11
US-08-486-013-59
Sequence 59, Application US/08486013
Patent No. 5731149
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,013
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:

LENGTH: 419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: misc_feature
LOCATION: 279..286
OTHER INFORMATION: /note= "N represents DNA that was
OTHER INFORMATION: not sequenced."
US-08-486-013-59

Query Match 2.1% Score 16; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTTCAGGTCCAGG 287
|||||
Db 114 TCCTTTCAGGTCCAGG 129

RESULT 12
US-08-482-279-59
Sequence 59, Application US/08482279
Patent No. 5840498
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,279
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: misc_feature
LOCATION: 279..286
OTHER INFORMATION: /note= "N represents DNA that was
OTHER INFORMATION: not sequenced."
US-08-482-279-59

Query Match 2.1%; Score 16; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

RESULT 13
US-08-342-268-59
; Sequence 59, Application US/08342268
; Patent No. 5844072
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/342,268
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P-UC 1205
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 1205
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 279..286
; OTHER INFORMATION: /note= "N represents DNA that was
; not sequenced."

US-08-342-268-59
Query Match 2.1%; Score 16; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

RESULT 14
US-09-015-968-59
; Sequence 59, Application US/09015968
; Patent No. 6057425
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE: 14-AUG-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,279
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 3003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 279..286
; OTHER INFORMATION: /note= "N represents DNA that was
; not sequenced."

US-09-015-968-59
Query Match 2.1%; Score 16; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

RESULT 15
US-09-397-386-59
; Sequence 59, Application US/09397386
; Patent No. 6300470
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods

US-09-397-386-59
Query Match 2.1%; Score 16; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

US-09-015-968-59
Query Match 2.1%; Score 16; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

RESULT 16
US-09-015-968-59
; Sequence 59, Application US/09015968
; Patent No. 6057425
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods

US-09-015-968-59
Query Match 2.1%; Score 16; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TCCTTCAGGTCGAGG 287
|||||
Db 114 TCCTTCAGGTCGAGG 129

```

1  TITLE OF INVENTION:  of Their Use
2  NUMBER OF SEQUENCES: 70
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: Campbell and Flore
5  STREET: 4370 La Jolla Village Drive, Suite 700
6  CITY: San Diego
7  STATE: California
8  COUNTRY: USA
9  ZIP: 92122
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/09/397,386
17 FILING DATE:
18 CLASSIFICATION:
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US/09/015,968
21 FILING DATE:
22 APPLICATION NUMBER: US 08/482,279
23 FILING DATE: 07-JUN-1995
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 08/342,268
26 FILING DATE: 18-NOV-1994
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 07/930,649
29 FILING DATE: 14-AUG-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 07/889,020
32 FILING DATE: 26-MAY-1992
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Campbell, Cathryn A.
35 REGISTRATION NUMBER: 31,815
36 REFERENCE/DOCKET NUMBER: P-UC 3003
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (619) 535-9001
39 TELEFAX: (619) 535-8949
40 INFORMATION FOR SEQ ID NO: 59:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 419 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: single
45 TOPOLOGY: linear
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: 279..286
49 OTHER INFORMATION: /note= "N represents DNA that was
50 OTHER INFORMATION: not sequenced."
51 US-09-397-386-59

```

```

Query Match      2.18; Score 16; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 272 TCCTTTCAGGTCACGG 287
      |||||

```

```

Db 114 TCCTTTCAGGTCACGG 129

```

```

Search completed: April 12, 2003, 09:03:11
Job time : 34.8447 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 22:52:34 ; Search time 1358.57 Seconds
(without alignments)
16087.603 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750
Perfect score: 751
Sequence: 1 ggaacaaatgccagaatctc.....gactgaactgtctgaaga 751

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	751	100.0	10562	6	E38420	E38420 Novel polyp
2	751	100.0	170121	9	AF064860	AF064860 Homo sapi
3	751	100.0	340000	9	HS21C080	AL163280 Homo sapi
4	649	86.4	1565	9	AB041413	AB041413 Homo sapi
5	649	86.4	1576	9	AB041416	AB041416 Homo sapi
6	517	68.8	933	9	AF145784	AF145784 Homo sapi
7	517	68.8	2762	9	AB020337	AB020337 Homo sapi
8	517	68.8	2775	6	E38419	E38419 Novel polyp
9	389	51.8	1360	9	AB041412	AB041412 Gorilla g
10	292	38.9	933	9	HS46078	AJ006078 Homo sapi
11	241	32.1	1570	9	AB041415	AB041415 Homo panis
12	176	23.4	1566	9	AB041414	AB041414 Pan trogl
13	138	18.4	1579	9	AB041417	AB041417 Pongo pyg
C 14	25	3.3	31	6	E38439	E38439 Novel polyp
C 15	24	3.2	31	6	E38440	E38440 Novel polyp
C 16	23	3.1	23	6	E38437	E38437 Novel polyp
C 17	23	3.1	927	10	AF254738	AF254738 Mus muscu
C 18	23	3.1	2309	9	HUMDGA	L77571 Homo sapien
C 19	23	3.1	39857	9	AC002522	AC002522 Homo sapi
C 20	23	3.1	43728	9	AC000395	AC000395 Homo sapi
C 21	23	3.1	191872	2	AC129837	AC129837 Papio cyn
C 22	23	3.1	196900	2	AC020851	AC020851 Mus muscu
C 23	22	2.9	47422	9	AL139817	AL139817 Human DNA
C 24	22	2.9	115597	9	AC010090	AC010090 Homo sapi
C 25	22	2.9	142893	9	AL360001	AL360001 Human DNA
C 26	22	2.9	215532	2	AC098478	AC098478 Homo sapi
C 27	22	2.9	226035	9	CNS01DVX	AL136293 Human chr
C 28	22	2.9	239057	9	AP001578	AP001578 Homo sapi
C 29	22	2.9	250529	9	HUAE000658	AE000658 Homo sapi
C 30	21	2.8	2021	9	AK000492	AK000492 Homo sapi
C 31	21	2.8	3220	9	AB067484	AB067484 Homo sapi
C 32	21	2.8	51803	9	AF176815	AF176815 Homo sapi
C 33	21	2.8	58882	2	RN107821	AL732854 Rattus no
C 34	21	2.8	62413	2	AC123612	AC123612 Mus muscu
C 35	21	2.8	64455	2	AC084134	AC084134 Homo sapi
C 36	21	2.8	73194	2	AC021266	AC021266 Homo sapi
C 37	21	2.8	77209	2	AC027366	AC027366 Homo sapi
C 38	21	2.8	77436	2	AC004581	AC004581 Homo sapi
C 39	21	2.8	79779	9	HSJ710H13	AL121712 Human DNA
C 40	21	2.8	81278	9	AC091062	AC091062 Homo sapi
C 41	21	2.8	86220	2	RN532D24	AL732655 Rattus no
C 42	21	2.8	92029	2	AL357492	AL357492 Homo sapi
C 43	21	2.8	94168	9	AL133415	AL133415 Human DNA
C 44	21	2.8	106172	9	AC004111	AC004111 Homo sapi
C 45	21	2.8	111182	9	AL359771	AL359771 Human DNA

ALIGNMENTS

RESULT 1
E38420
LOCUS E38420 Novel polypeptide.
DEFINITION E38420
ACCESSION E38420.1 GI:18626994
VERSION JP 2000245464-A/2.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 10562)
AUTHORS Narimatsu, H., Isshiki, S., Togayachi, A. and Sasaki, K.
TITLE Novel polypeptide
JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;

```
COMMENT
KYOWA HAKKO KOGYO CO LTD
OS Homo sapiens (human)
PN JP 2000245464-A/2
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
C12P21/02,
PC C12P21/08, C1201/68, G01N33/53//C12N1/21, C12R1:185, C12N5/10,
C12R1:91,
PC (C12P21/02, C12R1:185), (C12P21/02, C12R1:91), C12N15/00, C12N5/00,
C12N5/00, C12R1:91)
CC
FH Key Location/Qualifiers
FT promoter (1)..(5000)
FT exon (5001)..(5140)
FT exon (5001)..(5273)
FT exon (5459)..(5567)
FT exon (7427)..(7586)
FT exon (8234)..(10562).
FEATURES
Source
1..10562
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2610 a 2415 c 2574 g 2963 t
ORIGIN
Query Match 100.0%; Score 751; DB 6; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGACCAATGCCAGAACTCTGACGCTTTTATCTTACACCATGAAGTGACAGATGGTGC 60
Dbb 8000 GGACCAATGCCAGAACTCTGACGCTTTTATCTTACACCATGAAGTGACAGATGGTGC 8059
QY 61 AGATGTTAGACCTTTGTCTTAAGTGTAACTGTTAAACACACAGCAGCCGACTTCTGTATGAGC 120
Dbb 8060 AGATGTTAGACCTTTGTCTTAAGTGTAACTGTTAAACACACAGCAGCCGACTTCTGTATGAGC 8119
QY 121 GAGGTTCTAGAGTTCCAAACACGGGTCTCCTCCACACAGCTCAGCTCCTACATATAAA 180
Dbb 8120 GAGGTTCTAGAGTTCCAAACACGGGTCTCCTCCACACAGCTCAGCTCCTACATATAAA 8179
QY 181 CTAGACATCCTCATGCTTTTGAGGCTTAATCATGATTTGTTCTTCTTCAGATGGCT 240
Dbb 8180 CTAGACATCCTCATGCTTTTGAGGCTTAATCATGATTTGTTCTTCTTCAGATGGCT 8239
QY 241 TTCCGGAAGATGAGATTGATGATATATTGCTTCTGTTCTGGGGGCTCTTTTCTTTGTAT 300
Dbb 8240 TTCCGGAAGATGAGATTGATGATATATTGCTTCTGTTCTGGGGGCTCTTTTCTTTGTAT 8299
QY 301 TTAGCATGTACGCTTAATCTTCAAGACAGCTCCTTTGTTTACAGAAGACGGG 360
Dbb 8300 TTAGCATGTACGCTTAATCTTCAAGACAGCTCCTTTGTTTACAGAAGACGGG 8359
QY 361 AACTTCTTAAGCTCCAGATACAGACTGCAGGACAGACCTCCTTCTTCGCTCGCTGCTG 420
Dbb 8360 AACTTCTTAAGCTCCAGATACAGACTGCAGGACAGACCTCCTTCTTCGCTCGCTGCTG 8419
QY 421 GTGACCTATCCCAACAGATTGGCTGAGCGATGCGCATCGCCATCCGGCAGAGTGGGGAAA 480
Dbb 8420 GTGACCTATCCCAACAGATTGGCTGAGCGATGCGCATCGCCATCCGGCAGAGTGGGGAAA 8479
QY 481 GAGAGGATGGTGAAGGAAGACGCTCAAGACATCTCTCTCTCTGGGACCCACGACT 540
Dbb 8480 GAGAGGATGGTGAAGGAAGACGCTCAAGACATCTCTCTCTGGGACCCACGACT 8539
QY 541 GCAGCGAAAGAAAGAGGTGGACCGAGGAGCCAGCAGCAGCGGGACATATATCCAGAAG 600
Dbb 8540 GCAGCGAAAGAAAGAGGTGGACCGAGGAGCCAGCAGCAGCGGGACATATATCCAGAAG 8599
```

```
QY 601 GATTTCCTAGAGCTCTATTACAACTGACCTGAAGACCATGATGGGCATAGATGGTGC 660
Dbb 8600 GATTTCCTAGAGCTCTATTACAACTGACCTGAAGACCATGATGGGCATAGATGGTGC 8659
QY 661 CATCGCTTTTGTCCCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 720
Dbb 8660 CATCGCTTTTGTCCCTCAGCGCGCTTTGTGATGAAACAGACTCAGACATGTTTCATCAAT 8719
QY 721 GTTGACTATCTGACTGAAGTCTGCTTCTGAAGA 751
Dbb 8720 GTTGACTATCTGACTGAAGTCTGCTTCTGAAGA 8750
RESULT 2
LOCUS AF064860 170121 bp DNA linear PRI 05-MAR-2002
DEFINITION Homo sapiens chromosome 21 clone PAC 70124 map 21q22.3, complete
sequence.
ACCESSION AF064860
VERSION AF064860.2 GI:18958624
KEYWORDS HTG: HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170121)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kuch,J., Kawasaki,K., Asakawa,S.,
Shimani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichert,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Resselmann,L., Dagand,E., Haaf,T., Wehmeyer,S.,
Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
Reinhardt,R. and Yaspo,M.Laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
PUBMED 10830953
REFERENCE 2 (bases 1 to 170121)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 170121)
REFERENCE Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
AUTHORS Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
Rosenthal,A.
Direct Submission
Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 170121)
REFERENCE Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
AUTHORS Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
Direct Submission
Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Feb 27, 2002 this sequence version replaced gi:3171153.
FEATURES
Location/Qualifiers
source
1..170121
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC.70124"
BASE COUNT 45184 a 36756 c 37509 g 50672 t
ORIGIN
```

```
Query Match      100.0%; Score 751; DB 9; Length 170121;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACATGAAGTACAGATGCTGCG 60
Db 89814 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACATGAAGTACAGATGCTGCG 89873
QY 61 AGATGTTAGACCTTTGCTTAAGCTTTTAAACACACAGACACCCGACTTCTGTATGCAGC 120
Db 89874 AGATGTTAGACCTTTGCTTAAGCTTTTAAACACACAGACACCCGACTTCTGTATGCAGC 89933
QY 121 GAGTCTTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCTCTTACGATATAAA 180
Db 89934 GAGTCTTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCCTCTTACGATATAAA 89993
QY 181 CTAGACATCCCTCATGCTTTTTCAGTCTTAATCATTTGGATTGTTCTTTTCAGATGGCT 240
Db 89994 CTAGACATCCCTCATGCTTTTTCAGTCTTAATCATTTGGATTGTTCTTTTCAGATGGCT 90053
QY 241 TTCCGGAAGATGATGATGATATATTGCTTCTGTTCTGGGGCTCTTTGTTTGTAT 300
Db 90054 TTCCGGAAGATGATGATGATATATTGCTTCTGTTCTGGGGCTCTTTGTTTGTAT 90113
QY 301 TTTAGCATGTACAGTCTAATCTCTTCAAGAACAGCTCTCTTTTACAGAAAGACGGG 360
Db 90114 TTTAGCATGTACAGTCTAATCTCTTCAAGAACAGCTCTCTTTTACAGAAAGACGGG 90173
QY 361 AACTTCTCTTAAGTCTCCAGATACAGACTGCAGGCACACACCTCTCTCTCTCTCTGCTG 420
Db 90174 AACTTCTCTTAAGTCTCCAGATACAGACTGCAGGCACACACCTCTCTCTCTCTGCTG 90233
QY 421 GTGACCTATCCACAAACAGTTGGCTGAGCGCATGCCCATCCGCGACAGTGGGGGAAA 480
Db 90234 GTGACCTATCCACAAACAGTTGGCTGAGCGCATGCCCATCCGCGACAGTGGGGGAAA 90293
QY 481 GAGAGATGTGAGGGAAGAGAGCTGAAGACATTTCTCTCTCTGGGGACACACAGT 540
Db 90294 GAGAGATGTGAGGGAAGAGAGCTGAAGACATTTCTCTCTCTGGGGACACACAGT 90353
QY 541 GCAGCGGAAACGAAAGAGGTGGACACAGAGAGCCAGCGACGGGACATTTATCCAGAA 600
Db 90354 GCAGCGGAAACGAAAGAGGTGGACACAGAGAGCCAGCGACGGGACATTTATCCAGAA 90413
QY 601 GATTCTCTAGACGCTTATTACATCTGACCCCTGAGACACATGATGGGATAGATGGGTC 660
Db 90414 GATTCTCTAGACGCTTATTACATCTGACCCCTGAGACACATGATGGGATAGATGGGTC 90473
QY 661 CATCGCTTTTGCTCTCAGCGCGGCTTTGTGATCAAAACAGACTCAGACATGTTTCATCAAT 720
Db 90474 CATCGCTTTTGCTCTCAGCGCGGCTTTGTGATCAAAACAGACTCAGACATGTTTCATCAAT 90533
QY 721 GTTGACTATCTGACTGAAGTCTTCTGGAAGA 751
Db 90534 GTTGACTATCTGACTGAAGTCTTCTGGAAGA 90564

RESULT 3
LOCUS HS21C080 340000 bp DNA linear PRI 24-MAY-2000
DEFINITION Homo sapiens chromosome 21 segment HS21C080.
ACCESSION AL163280 AP001735 BA000005
VERSION AL163280.2 GI:7717369
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
```

```
TITLE
JOURNAL

COMMENT
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gsic.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: Shimizu@dm-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
Innestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1.125346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_lib="P160L9, 5' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121897"
73410..243533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF064860"
224137..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11.3-5 PAC library"
/note="Accession No. AF121782"
286628..340000
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088..1566
repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398..2495
repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
complement(2496..2878)
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
2879..2931
repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2932..3004
repeat_region
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3005..3074
repeat_region
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3473..19157
gene
/gene="SH3BGR"
join(<3473..3565,12649..12678,15331..15460,18739..19157)
mRNA
/gene="SH3BGR"
join(3473..3565,12649..12678,15331..15426)
CDS
/partial
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/protein_id="CAB90445.1"
/db_xref="GI:7717370"
/translation="GSEKAEGETEAKGSGEDVGNLPPAQEKNEEGETATETEE"
IAMGAGSEAEFEETAEAGEPEDEDS"
3473..3565
exon
/gene="SH3BGR"
/number=4
5542..5829
repeat_region
/note="AluSX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(6887..7067)
/note="L1MD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region
complement(8482..8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(8865..8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region
9131..9424
/note="AluSG"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(9723..9882)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region
complement(9889..10077)
/note="L2"
```

```
repeat_region
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10078..10429)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(10430..10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
10980..11000
repeat_region
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region
complement(11002..11282)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
exon
/gene="SH3BGR"
/number=5
complement(13115..13425)
/note="AluSX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
repeat_region
/note="(CA)n"
/rpt_family="Simple_repeat"
```

```
Query Match 100.08; Score 751; DB 9; Length 340000;
Best Local Similarity 100.08; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGGC 60
|||||
Db 163222 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTACAGATGCTGGC 163281

QY 61 AGATGTTAGACCTTTGTGCTTAACCTGTTAAACACACAGACCGGACTTCTGTATGAGC 120
|||||
Db 163282 AGATGTTAGACCTTTGTGCTTAACCTGTTAAACACACAGACCGGACTTCTGTATGAGC 163341

QY 121 GAGGTTCTAGAGTTTCCAAACACAGGGTCTCTCTCCACCTCAGCCTCTCAGCATATAA 180
|||||
Db 163342 GAGGTTCTAGAGTTTCCAAACACAGGGTCTCTCTCCACCTCAGCCTCTCAGCATATAA 163401

QY 181 CTAGACACATCCTCATGCTTTTGAGGCTTAATCATTTGATTTTCTCTTTCAGATGGCT 240
|||||
Db 163402 CTAGACACATCCTCATGCTTTTGAGGCTTAATCATTTGATTTTCTCTTTCAGATGGCT 163461

QY 241 TTCCCGAAGATGAGATTGATGATATTTGCTTCTGTTTCGCGGGGCTCTTTGTTTGTAT 300
|||||
Db 163462 TTCCCGAAGATGAGATTGATGATATTTGCTTCTGTTTCGCGGGGCTCTTTGTTTGTAT 163521

QY 301 TTAGCATGTACAGTCTAAATCCTTTCAAGAACAGTCCTTTGTTTCAAGAAAGACGGG 360
|||||
Db 163522 TTAGCATGTACAGTCTAAATCCTTTCAAGAACAGTCCTTTGTTTCAAGAAAGACGGG 163581

QY 361 AACTTCCTTAAGTCCCGAGATACAGATGCGAGGACACCTCCCTTCCTCGCTGCTG 420
|||||
Db 163582 AACTTCCTTAAGTCCCGAGATACAGATGCGAGGACACCTCCCTTCCTCGCTGCTG 163641
```


QY 421 GTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCGGCAGACCTGGGGGAAA 480
|||||
Db 163642 GTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCGGCAGACCTGGGGGAAA 163701
|||||
QY 481 GAGAGGATGGTGAAGGAAAGAGCAGCTGAAGACATTTTCCTCTGGGGACCAACAGCAGT 540
|||||
Db 163702 GAGAGGATGGTGAAGGAAAGAGCAGCTGAAGACATTTTCCTCTGGGGACCAACAGCAGT 163761
|||||
QY 541 GCAGCGGAACGAAAGAGCTGGACAGAGAGCCAGCCAGCCAGCGGGGACATATCCAGAG 600
|||||
Db 163762 GCAGCGGAACGAAAGAGCTGGACAGAGAGCCAGCCAGCCAGCGGGGACATATCCAGAG 163821
|||||
QY 601 GATTTCCTAGACCTCTATTACAACTGACCCCTGAAGACCATGATGGCATAGATGGGTC 660
|||||
Db 163822 GATTTCCTAGACCTCTATTACAACTGACCCCTGAAGACCATGATGGCATAGATGGGTC 163881
|||||
QY 661 CATCGCTTTTGCTCAGCGCGGCTTTGTGATGAAGACAGCTCAGACATGTTCAATCAAT 720
|||||
Db 163882 CATCGCTTTTGCTCAGCGCGGCTTTGTGATGAAGACAGCTCAGACATGTTCAATCAAT 163941
|||||
QY 721 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
|||||
Db 163942 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 163972
|||||

RESULT 4
AB041413
LOCUS
DEFINITION
Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc
beta1,3-galactosyltransferase 5, partial cds.
ACCESSION
AB041413
VERSION
AB041413.1 GI:7593020
KEYWORDS
Homo sapiens (isolate:human-NR) DNA.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Li,Y. and Saitou,N.
TITLE
Silver Project
JOURNAL
Published Only in Database (2000)
REFERENCE
2 (bases 1 to 1565)
AUTHORS
Li,Y. and Saitou,N.
TITLE
Direct Submission
JOURNAL
Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL:http://sayer.lab.nig.ac.jp/~silver/ Tel:81-559-81-6790,
Fax:81-559-81-6789)
FEATURES
Location/Qualifiers
1..1565
/organism="Homo sapiens"
/isolate="human-NR"
/db_xref="taxon:9606"
/note="Human sequence used for primer design based on Acc#
AB020337"
exon
1..27
intron
28..674
gene
675..1565
CDS
675..>1565
/gene="beta1,3-GalT 5"
/codon_start=1
/product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
/protein_id="BA094498.1"
/db_xref="gi:7593021"
/translation="MAPPKRLMWICLLVLGALCLVFSMYSLNPFKEQSFVYKKDGNF
LKLPTDOROTPPPLVLLVSSHKQLAERNAIRQWGRKERYXVKGLKTFLLGTTSS
AAETKVDQSRHGDIIQKDELVDVYNULTKMGIWHRFCPPAARVMTKTDMSF
INVDYELLKKNRTRFTFTGLKNEFFIRQFFSKWFSKSEYFWDYRPPFCSGTG
YVSGDVASQYINVSVPYIKLEDVFGVGLERLNIRLELSQPTFTFPGLRFSVC

exon
LPERRIVACHFIKPRTLWTGR*
675..>1565
/gene="beta1,3-GalT 5"
/number=4
BASE COUNT 410 a 370 c 385 g 398 t 2 others
ORIGIN
Query Match 86.4%; Score 649; DB 9; Length 1565;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 749; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGACCAATGCCAGAAATCTCTGAGCTTTTATCTTACACATGAAGTACAGATGCTGCG 60
|||||
Db 441 GGACCAATGCCAGAAATCTCTGAGCTTTTATCTTACACATGAAGTACAGATGCTGCG 500
|||||
QY 61 AGATGTTAGACCTTTGTGCTTAATCTTTAAACACACAGACCCCGACTTCTCTATGACAG 120
|||||
Db 501 AGATGTTAGACCTTTGTGCTTAATCTTTAAACACACAGACCCCGACTTCTCTATGACAG 560
|||||
QY 121 GAGTTCTAGAGTTTCCAAACACAGGCTCTCTCTCCACCTCAGCTCCTTAGCATAAA 180
|||||
Db 561 GAGTTCTAGAGTTTCCAAACACAGGCTCTCTCTCCACCTCAGCTCCTTAGCATAAA 620
|||||
QY 181 CTAGACACATCCTCATGCTTTTGGAGTCTAATCAATTTGGATTTTCTTCCATTCAGATGGCT 240
|||||
Db 621 CTAGACACATCCTCATGCTTTTGGAGTCTAATCAATTTGGATTTTCTTCCATTCAGATGGCT 680
|||||
QY 241 TTCCCGAAGATGAGATGATATATTTGCTTCTGGTCTCTGGGGCTCTTTGTTTGTAT 300
|||||
Db 681 TTCCCGAAGATGAGATGATATATTTGCTTCTGGTCTCTGGGGCTCTTTGTTTGTAT 740
|||||
QY 301 TTTAGCATGTACAGTCTAAATCCTTTCAAGAACAGTCTCTTTTACAAAGAACACGGG 360
|||||
Db 741 TTTAGCATGTACAGTCTAAATCCTTTCAAGAACAGTCTCTTTTACAAAGAACACGGG 800
|||||
QY 361 AACTTCTTAAGCTCCAGATACAGACTCAGCGACAGACCTCCCTTCTCTGCTGCTG 420
|||||
Db 801 AACTTCTTAAGCTCCAGATACAGACTCAGCGACAGACCTCCCTTCTCTGCTGCTG 860
|||||
QY 421 GTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCGGCAGACCTGGGGGAAA 480
|||||
Db 861 GTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCGGCAGACCTGGGGGAAA 920
|||||
QY 481 GAGAGGATGGTGAAGGAAAGAGCAGCTGAAGACATTTCTCTCTGGGACCAACAGCAGT 540
|||||
Db 921 GAGAGGATGGTGAAGGAAAGAGCAGCTGAAGACATTTCTCTCTGGGACCAACAGCAGT 980
|||||
QY 541 GCAGCGGAACGAAAGAGCTGGACAGAGAGCCAGCGACAGCGGGGACATATCCAGAG 600
|||||
Db 981 GCAGCGGAACGAAAGAGCTGGACAGAGAGCCAGCGACAGCGGGGACATATCCAGAG 1040
|||||
QY 601 GATTTCCTAGACCTCTATTACAACTGACCCCTGAAGACCATGATGGCATAGATGGGTC 660
|||||
Db 1041 GATTTCCTAGACCTCTATTACAACTGACCCCTGAAGACCATGATGGCATAGATGGGTC 1100
|||||
QY 661 CATCGCTTTTGCTCAGCGCGGCTTTGTGATGAAGACAGCTCAGACATGTTCAATCAAT 720
|||||
Db 1101 CATCGCTTTTGCTCAGCGCGGCTTTGTGATGAAGACAGCTCAGACATGTTCAATCAAT 1160
|||||
QY 721 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
|||||
Db 1161 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 1191
|||||
RESULT 5
AB041416
LOCUS
DEFINITION
Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc
beta1,3-galactosyltransferase 5, partial cds.
ACCESSION
AB041416
VERSION
AB041416.1 GI:7593026
KEYWORDS
Homo sapiens (isolate:#056) DNA.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Liu, Y. and Saitou, N.
TITLE Silver Project
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 1576)
AUTHORS Liu, Y. and Saitou, N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
URL: <http://saver.lab.nig.ac.jp/~silver/>, Tel: 81-559-81-6790,
Fax: 81-559-81-6789)
FEATURES
source
Location/Qualifiers
1..1576
/organism="Homo sapiens"
/isolate="#056"
/db_xref="taxon:9606"
/note="human sequence used for primer design based on Acc#
AB020337"
<1..27
/number=3
28..674
/number=3
675..1576
/gene="betal.3-Galt 5"
675..>1576
/gene="betal.3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BAA94501.1"
/db_xref="GI:7593027"
/translation="MAPPKRLMYICLLVLGALCLVFSYSLNPFQSFVYKKGDNF
LKLPDTCROTPTPELVLLVSSHQKLAERMAIQTWKSEYTKGKGLKFFLLGTSS
AAETKEYDQSRHGDIQKDFLVNLTLMGIEWHRFCPPQAAPVMTKDSDF
INVYIELLLKRNTRTFITGFLNEPFIQPFKFWFSKSEYFMDRPPFCSGT
LVFSGDVASQYKNSVPYKLEDFVFGCLERLNIRLELSQPTFFPPGLRFSVC
LFRIVACHIKPTLLDYQWALE"
675..>1576
/gene="betal.3-Galt 5"
/number=4
415 a 373 c 387 g 401 t
BASE COUNT 415 a 373 c 387 g 401 t
ORIGIN
Query Match 86.4%; Score 549; DB 9; Length 1576;
Best Local Similarity 93.7%; Pred. No. 0;
Matches 749; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCACCAATGCCAGAACTCTGAGCTTTTATCTACACCATGAAGTCAGATGCTGGC 60
Db 441 GGACCAATGCCAGAACTCTGAGCTTTTATCTACACCATGAAGTCAGATGCTGGC 500
QY 61 AGATGTTAGACCTTGTGCTTAACTGTTTAAACACACAGCAGCCGACTTCTGTATGAGC 120
Db 501 AGATGTTAGACCTTGTGCTTAACTGTTTAAACACACAGCAGCCGACTTCTGTATGAGC 560
QY 121 GAGGTTCTAGAGTTCCAAACACGGGTCTCTCTCCACCTCAGCTCTCAGATATAA 180
Db 561 GAGGTTCTAGAGTTCCAAACACGGGTCTCTCTCCACCTCAGCTCTCAGATATAA 620
QY 181 CTAGACATCCTCATGCTTTTGGGCTTAATCATTTGATTTTCTCTTTCAGATGGCT 240
Db 621 CTAGACATCCTCATGCTTTTGGGCTTAATCATTTGATTTTCTCTTTCAGATGGCT 680
QY 241 TTCGCGAAGATGAGATTGATGATATTTGCTTCTGTTCTGGGGCTCTTTTGTAT 300
Db 681 TTCGCGAAGATGAGATTGATGATATTTGCTTCTGTTCTGGGGCTCTTTTGTAT 740
QY 301 TTACATGTCAGTCTAAATCTTTCAAGAAGCAGTCTTTGTTTACAGAAGACGGG 360
Db 741 TTACATGTCAGTCTAAATCTTTCAAGAAGCAGTCTTTGTTTACAGAAGACGGG 800

QY 361 AACTTCCTTAAGCTCCAGATACAGACTGCAGGAGACACCTCCCTTCTCGTCTGCTG 420
Db 801 AACTTCCTTAAGCTCCAGATACAGACTGCAGGAGACACCTCCCTTCTCGTCTGCTG 860
QY 421 GTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAA 480
Db 861 GTGACCTCATCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAA 920
QY 481 GAGAGGATGGTGAAGGGAAGCAGCTGAAGACATTTCTCTCTGGGACACACAGCAGT 540
Db 921 GAGAGGACGGTGAAGGGAAGCAGCTGAAGACATTTCTCTCTGGGACACACAGCAGT 980
QY 541 GCAGCGGAACCAAGAGAGTGGACAGGAGGAGCCAGCCAGCGGACATATCCAGAG 600
Db 981 GCAGCGGAACCAAGAGAGTGGACAGGAGGAGCCAGCGGACATATCCAGAG 1040
QY 601 GATTTCTAGAGCTCTATTACAAATCTGACCTCTGAAGACCATGATGGGCATAGATGGTGC 660
Db 1041 GATTTCTAGAGCTCTATTACAAATCTGACCTCTGAAGACCATGATGGGCATAGATGGTGC 1100
QY 661 CATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAGAACAGACTCAGACATGTTTCAAT 720
Db 1101 CATCGCTTTTGTCTCAGCGCGCTTTGTGATGAAGAACAGACTCAGACATGTTTCAAT 1160
QY 721 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 751
Db 1161 GTTGACTATCTGACTGAACCTGCTTCTGAAGA 1191
RESULT 6
AF145784 933 bp DNA linear PRI 14-NOV-1999
LOCUS Homo sapiens betal,3 galactosyltransferase-v (B3GALT5) gene,
complete cds.
ACCESSION AF145784
VERSION AF145784.1 GI:6409192
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Zhou, D., and Hennet, T.
TITLE Molecular cloning of a human UDP-galactose:GlcNAc betal,3GalNAc
betal,3 galactosyltransferase gene encoding an O-linked
core3-elongation enzyme
JOURNAL Eur J Biochem. 263 (2), 571-576 (1999)
MEDLINE 99337698
PUBMED 10406968
REFERENCE 2 (bases 1 to 933)
AUTHORS Zhou, D., and Hennet, T.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1999) Physiology, University of Zurich,
Winterthurerstrasse 190, Zurich 8057, Switzerland
FEATURES
source
Location/Qualifiers
1..933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22"
1..933
/gene="B3GALT5"
1..933
/gene="B3GALT5"
/note="glycosyltransferase"
/codon_start=1
/product="betal,3 galactosyltransferase-v"
/protein_id="AA07880.1"
/db_xref="GI:6409193"
/translation="MAPPKRLMYICLLVLGALCLVFSYSLNPFQSFVYKKGDNF
LKLPDTCROTPTPELVLLVSSHQKLAERMAIQTWKSEYTKGKGLKFFLLGTSS
AAETKEYDQSRHGDIQKDFLVNLTLMGIEWHRFCPPQAAPVMTKDSDF

INVLYTELLKKNRTRFTFTGFLKLNFFPIQPPSKWFKSVKSEYPWDRVPPFCSTG
YVSGDVASOVNVKSVYKLEDFVGLCLRLNIRLELHNSQPTFFEGGLRFSVC
LRRIVACHFIKPRLLDYWALENSRGECCPPV

BASE COUNT 227 a 241 g 233 t
ORIGIN
Query Match 68.8%; Score 517; DB 9; Length 933;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 ATGGCTTCCCGAAGATGAGATGATATATTGCTTCTGGTCTGGGGGCTTTTGT 294
DB 1 ATGGCTTCCCGAAGATGAGATGATATATTGCTTCTGGTCTGGGGGCTTTTGT 60
QY 295 TTGTATTATTAGCATGACAGTCTAAATCCCTTTCAAAAGACAGTCTTTTGTACAGAAA 354
DB 61 TTGTATTATTAGCATGACAGTCTAAATCCCTTTCAAAAGACAGTCTTTTGTACAGAAA 120
QY 355 GACGGGAACCTTCTTAAGTCCCGATACAGACTGACGACAGACACCTTCTCTGTC 414
DB 121 GACGGGAACCTTCTTAAGTCCCGATACAGACTGACGACAGACACCTTCTCTGTC 180
QY 415 CTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCCGGACAGCTGG 474
DB 181 CTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCCGGACAGCTGG 240
QY 475 GGGAAAGAGAGGATGGTGAAGGAAGACAGCTGAAGACATTTCTCTCTGGGGACCA 534
DB 241 GGGAAAGAGAGGATGGTGAAGGAAGACAGCTGAAGACATTTCTCTCTGGGGACCA 300
QY 535 AGCAGTGCAGGGGAAACGAAGAGGTGGACAGGAGAGGACGACGACGGGACATATC 594
DB 301 AGCAGTGCAGGGGAAACGAAGAGGTGGACAGGAGAGGACGACGACGGGACATATC 360
QY 595 CAGAGGATTTCTAGACGCTTATTACATCTGACCTCGAAGACCATGATGGGCATAGAA 654
DB 361 CAGAGGATTTCTAGACGCTTATTACATCTGACCTCGAAGACCATGATGGGCATAGAA 420
QY 655 TGGTCCATCGCTTTCTCTGAGCGGCGTTTGTGATGAAACACAGCTGACATGTC 714
DB 421 TGGTCCATCGCTTTCTCTGAGCGGCGTTTGTGATGAAACACAGCTGACATGTC 480
QY 715 ATCAATGTGACTATCTGACTGACTGCTTCTGAGA 751
DB 481 ATCAATGTGACTATCTGACTGACTGCTTCTGAGA 517

RESULT 7
AB020337
LOCUS
DEFINITION Homo sapiens mRNA for UDP-Gal:GlcNAc betal,3-galactosyltransferase 5, complete cds.
ACCESSION AB020337
VERSION AB020337.1 GI:4835502
KEYWORDS UDP-Gal:GlcNAc betal,3-galactosyltransferase 5.
SOURCE Homo sapiens Adenocarcinoma cell_line:Colo 205 cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Ishiki,S., Togayachi,A., Kudo,T., Nishihara,S., Watanabe,M., Kubota,T., Kitajima,M., Shiraishi,N., Sasaki,K., Andoh,T. and Narimatsu,H.
TITLE Cloning, expression, and characterization of a novel UDP-galactose:beta-N-acetylglucosamine betal,3-galactosyltransferase (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived therefrom
J. Biol. Chem. 274 (18), 12499-12507 (1999)
JOURNAL 9230269
MEDLINE 2 (bases 1 to 2762)
AUTHORS Ishiki,S., Togayachi,A. and Narimatsu,H.
TITLE Direct Submission

JOURNAL Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University, Institute of Life Science; 1-236, Tangi-cho, Hachioji, Tokyo 192-8577, Japan (E-mail: sisshiki@po.iiinet.or.jp, Tel.81-426-91-9466, Fax:81-426-91-9315)

FEATURES
source
1. .2762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/cell_line="Colo 205"
/cell_type="Adenocarcinoma"
1. .2762
/gene="betal,3-GalT 5"
1. .273
/gene="betal,3-GalT 5"
/number=1
274. .433
/gene="betal,3-GalT 5"
/number=3
434. .2762
/gene="betal,3-GalT 5"
/number=4
434. .1366
/gene="betal,3-GalT 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BAA77664.1"
/db_xref="GI:4835503"
/translation="MAFFKRLMLVLCILVGLALCLYFSMYSLNPPFKSQSFYKKDGNF
LKLPTDCRQTPFFLVLLVTSKQLAERAIROTWGRKMRVKGKQLKTFLLGTTSS
AAETKVDQESQHGDIQKDFLVYVNLTKTMGTEWVHRFCPOAFAFVKMTSDMF
INVDYLTLLKKNRTFFELFKLNEPFIROPSPKHFYSKSEYPWDRVPPFCSTG
YVSGDVASOVNVKSVYKLEDFVGLCLRLNIRLELHNSQPTFFEGGLRFSVC
LRRIVACHFIKPRLLDYWALENSRGECCPPV"
polyA_site
2762
/gene="betal,3-GalT 5"
/note="45 a nucleotides"
BASE COUNT 651 a 703 c 679 g 729 t
ORIGIN

Query Match 68.8%; Score 517; DB 9; Length 2762;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 235 ATGGCTTCCCGAAGATGAGATGATATATTGCTTCTGGTCTGGGGGCTTTTGT 294
DB 434 ATGGCTTCCCGAAGATGAGATGATATATTGCTTCTGGTCTGGGGGCTTTTGT 493
QY 295 TTGTATTATTAGCATGACAGTCTAAATCCCTTTCAAAAGACAGTCTTTTGTACAGAAA 354
DB 494 TTGTATTATTAGCATGACAGTCTAAATCCCTTTCAAAAGACAGTCTTTTGTACAGAAA 553
QY 355 GACGGGAACCTTCTTAAGTCCCGATACAGACTGACGACAGACACCTTCTCTGTC 414
DB 554 GACGGGAACCTTCTTAAGTCCCGATACAGACTGACGACAGACACCTTCTCTGTC 613
QY 415 CTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCCGGACAGCTGG 474
DB 614 CTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGCCATCCCGGACAGCTGG 673
QY 475 GGGAAAGAGAGGATGGTGAAGGAAGACAGCTGAAGACATTTCTCTCTGGGGACCA 534
DB 674 GGGAAAGAGAGGATGGTGAAGGAAGACAGCTGAAGACATTTCTCTCTGGGGACCA 733
QY 535 AGCAGTGCAGGGGAAACGAAGAGGTGGACAGGAGAGGACGACGACGGGACATATC 594
DB 734 AGCAGTGCAGGGGAAACGAAGAGGTGGACAGGAGAGGACGACGACGGGACATATC 793
QY 595 CAGAGGATTTCTAGACGCTTATTACATCTGACCTCGAAGACCATGATGGGCATAGAA 654
DB 794 CAGAGGATTTCTAGACGCTTATTACATCTGACCTCGAAGACCATGATGGGCATAGAA 853

Qy	655	TGGTCCATCGCTTTTGTCTCAGCGGGTGTGTGATGAAGAAACAGACTCAGACATGTC	714
Db <td>854</td> <td>TGGGTCCATCGCTTTTGTCTCAGCGGGTGTGTGATGAAGAAACAGACTCAGACATGTC</td> <td>913</td>	854	TGGGTCCATCGCTTTTGTCTCAGCGGGTGTGTGATGAAGAAACAGACTCAGACATGTC	913
Qy <td>715</td> <td>ATCATGTGTGACTATCTGACTGAACTGCTCTCTGAAGA</td> <td>751</td>	715	ATCATGTGTGACTATCTGACTGAACTGCTCTCTGAAGA	751
Db <td>914</td> <td>ATCAATGTGTACTATCTGACTGAACTGCTCTCTGAAGA</td> <td>950</td>	914	ATCAATGTGTACTATCTGACTGAACTGCTCTCTGAAGA	950
RESULT	B		
E38419			
LOCUS	E38419	2775 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Novel polypeptide.		
ACCESSION	E38419		
VERSION	E38419.1	GI:18626993	
KEYWORDS	JP 2000245464-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2775)		
TITLE	Narimatsu,H., Isshiki,S., Togayauchi,A. and Sasaki,K.		
JOURNAL	Novel polypeptide		
COMMENT	Patent: JP 2000245464-A 1 12-SEP-2000; KYOWA HAKKO KOGYO CO LTD OS Homo sapiens (human) PN JP 2000245464-A/1 PD 12-SEP-2000 PR 25-FEB-1999 JP 1999047571		
PR	HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI		
PC	KATSUTOSHI SASAKI		
PC	C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC		
PC	C12P21/02,		
PC	C12P21/08, C12P1/68, G01N33/53, C12N1/21, C12R1/185, C12N5/10,		
PC	C12R1/91),		
PC	(C12P21/02, C12R1/185), (C12P21/02, C12R1/91), C12N15/00, C12N5/00,		
PC	(C12N5/00, C12R1/91)		
CC			
FF	Key	Location/Qualifiers	
FF	CDS	(402)..(1331).	
FEATURES	1..2775		
source			
BASE COUNT	681 a 698 c 669 g 727 t		
ORIGIN			
Query Match	68.8%	Score 517; DB 6; Length 2775;	
Best Local Similarity	100.0%	Pred. No. 1.1e-303;	
Matches 517; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	235	ATGGCTTTCCGAGATGAGATTGATGTATTTGGCTTCTGGTCTGGGGGCTCTTTGT	294
Db	402	ATGGCTTTCCGAGATGAGATTGATGTATTTGGCTTCTGGTCTGGGGGCTCTTTGT	461
QY	295	TTGTATTTTAGCATGTACAGTCTAAATCCCTTTCAAAGAACAGTCTTTTACAGAAA	354
Db	462	TTGTATTTTAGCATGTACAGTCTAAATCCCTTTCAAAGAACAGTCTTTTACAGAAA	521
QY	355	GACGGGAATTCCTTAAGCTCCAGATACAGATGGAGCAGACATCCCTCTCTCTGTC	414
Db	522	GACGGGAATTCCTTAAGCTCCAGATACAGATGGAGCAGACATCCCTCTCTCTGTC	581
QY	415	CTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG	474
Db	582	CTGCTGGTGACCTCATCCCAAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG	641
QY	475	GGGAAGAGAGATGGTGAAGGGAAGACACTGAAGACATCTTCTCTCTGGGGACCACC	534
Db	642	GGGAAGAGAGATGGTGAAGGGAAGACACTGAAGACATCTTCTCTCTGGGGACCACC	701
QY	535	AGCAGTGCAGCGGAACAAAGAGGTTGGACACAGAGACCCAGGACACCGGGGACATATC	594

Db	702	ACGAGTGCACGGGAACCAAGAGAGTGGACACGAGGAGCCACCGGACGACATTATC	761
QY	595	CAGAAGAGATTCTCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA	654
Db	762	CAGAAGAGATTCTCTAGACGCTCTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA	821
QY	655	TGGGTCCATCCCTTTCTCTCAGCGGGGCTTTGTGATGAAACAGACTCAGACATGTC	714
Db	822	TGGGTCCATCCCTTTCTCTCAGCGGGGCTTTGTGATGAAACAGACTCAGACATGTC	881
QY	715	ATCAATGTTGACTATCTGACTGAAGTCTCTCTGAGA	751
Db	882	ATCAATGTTGACTATCTGACTGAAGTCTCTCTGAGA	918
RESULT 9	AB041412	AB041412	1360 bp DNA linear PRI 17-OCT-2000
LOCUS		Gorilla gorilla betal,3-Galt 5 gene for UDP-Gal:GlcNAc	
DEFINITION		betal,3-galactosyltransferase 5, partial cds.	
ACCESSION		AB041412	
VERSION		AB041412.1	GI:7593018
KEYWORDS			
SOURCE		Gorilla gorilla (isolate:#085) DNA.	
ORGANISM		Gorilla gorilla	
REFERENCE		Eukaryota; Metazoa	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.	
TITLE		1 (bases 1 to 1360)	
JOURNAL		Liu, Y. and Saitou, N.	
REFERENCE		Silver Project	
AUTHORS		Published Only in DataBase (2000)	
TITLE		2 (bases 1 to 1360)	
JOURNAL		Liu, Y. and Saitou, N.	
REFERENCE		Direct Submission	
AUTHORS		Submitted (11-APR-2000) Naruya Saitou, National Institute of	
TITLE		Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,	
JOURNAL		Shizuoka 411-8540, Japan (E-mail:nsaitou@genetics.nig.ac.jp,	
REFERENCE		URL: http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,	
AUTHORS		Fax:81-559-81-6789)	
TITLE		Location/Qualifiers	
JOURNAL		1. .1360	
FEATURES		/organism="Gorilla gorilla"	
source		/isolate="#085"	
		/db_xref="taxon:9593"	
		/note="human sequence used for primer design based on Acc#	
		AB020337"	
		<1. .466	
		/number=3	
		467. .1360	
		/gene="betal.3-Galt 5"	
		467. >1360	
		/gene="betal.3-Galt 5"	
		/codon_start=1	
		/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"	
		/protein_id="BAA94497.1"	
		/db_xref="GI:7593019"	
		/translation="MAPPPKRLMLICLLVGLCALCYFLSYSLNPNFKQSIVYKKDWNF	
		LKLPDCTQTPPEFLVLLVTSRHKQLAERMAIRQTMGKRTVKGQKLTFFLLGTTSS	
		AATYKVDQDSRRHGDIOKDFLDVYNTLTKNMGIEWHRFCQAAFPVKMTDSGF	
		INDVLTLLKKNRTTREFGTFLNREPIRQPSKWFVSKSYPDWRVPPFCDSGT	
		YVPSGDVASOVYVSESPYIKLEDVFVGLCLERLINLELHSQPTFFPGGLRFVSC	
		RRFRIVACHIRKPTLLDVAQA"	
		467. .>1360	
		/gene="betal.3-Galt 5"	
		/number=4	
exon			
BASE COUNT	334 a	331 c	341 g 351 t 3 others
ORIGIN			

Query Match 51.8%; Score 389; DB 9; Length 1360;
Best Local Similarity 99.1%; Pred. No. 1.8e-225;
Matches 739; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 AATGCCAGAACTCTGAGCCTTTTATCTTACACCATGAAGTACAGATGCTGCAGATG 65
 Db 238 AATGCCAGAACTCTGAGCCTTTTATCTTACACCATGAAGTACAGATGCTGCAGATG 297
 QY 66 TTAGACTTTTGCTTAACTGTTTAAACACACAGACCCGACTTCTGTATCAGCGAGGT 125
 Db 298 TTAGACTTTTGCTTAACTGTTTAAACACACAGACCCGACTTCTGTATCAGCGAGGT 357
 QY 126 TCTAGAGTTTCCAAAACAGCGGTCTCCTCTCCACCTCAGCCTCTAGCATATAAACTAGA 185
 Db 358 TCTAGAGTTTCCAAAACAGCGGTCTCCTCTCCACCTCAGCCTCTAGCATATAAACTAGA 417
 QY 186 CATCTCCTCATGCTTTTGAAGTCTTAATCATTTGTTTCTCTTTCAGATGGCTTTCCC 245
 Db 418 CATCTCCTCATGCTTTTGAAGTCTTAATCATTTGTTTCTCTTTCAGATGGCTTTCCC 477
 QY 246 GAAGATGAGATTGATGATATATTGCTTCTGCTTCTGGGGCTCTTTGTTTATTTAG 305
 Db 478 GAAGATGAGATTGATGATATATTGCTTCTGCTTCTGGGGCTCTTTGTTTATTTAG 537
 QY 306 CATGTACAGTTAAATCCTTCAAGAACAGTCCCTTTTACAGAAAGACGGGAAGCTT 365
 Db 538 CATGTACAGTTAAATCCTTCAAGAACAGTCCCTTTTACAGAAAGACGGGAAGCTT 597
 QY 366 CCTTAAGCTCCAGATACAGACTGCAGGACAGACCTCCTCTCTGCTGCTGGTAC 425
 Db 598 CCTTAAGCTCCAGATACAGACTGCAGGACAGACCTCCTCTCTGCTGCTGGTAC 657
 QY 426 CTATCCCAACAGTTCGCTGAGCGCATGCCATCCGACAGACGCTGGGGAAGAGAG 485
 Db 658 CTATCCCAACAGTTCGCTGAGCGCATGCCATCCGACAGACGCTGGGGAAGAGAG 717
 QY 486 GATGGTGAAGGGAAGACAGTGAAGACATTTCTCTCTGGGACACACAGCAGTGCAGC 545
 Db 718 GACAGTGAAGGGAAGACAGTGAAGACATTTCTCTCTGGGACACACAGCAGTGCAGC 777
 QY 546 GGAACCAAGAGTGGACACAGAGACGACGACGACGCGGACATATCCAGAGGATTT 605
 Db 778 GGAACCAAGAGTGGACACAGAGACGCGGACGACGCGGACATATCCAGAGGATTT 837
 QY 606 CCTAGAGCTCTATTACAATCTGACCTGAAGACCATGATGGGCATAGATGGTCCATCG 665
 Db 838 CCTAGAGCTCTATTACAATCTGACCTGAAGACCATGATGGGCATAGATGGTCCATCG 897
 QY 666 CTTTGTCTCAGCGCGCTTTGTGATGAACAGACTCAGACATGTCATCAATGTTGA 725
 Db 898 CTTTGTCTCAGCGCGCTTTGTGATGAACAGACTCAGACATGTCATCAATGTTGA 957
 QY 726 CTATCTGACTGAAGTCTCTGAAGA 751
 Db 958 CTATCTGACTGAAGTCTCTGAAGA 983

RESULT 10
 HSA6078
 LOCUS Homo sapiens beta3gal-T6 gene. 933 bp DNA linear PRI 11-MAY-2000
 DEFINITION
 ACCESSION A7006078
 VERSION A7006078.1 GI:7799922
 KEYWORDS beta-1,3-galactosyltransferase; beta3gal-T6 gene.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Amado, M., Carneiro, F. and Clausen, H.
 TITLE Cloning and expression of two beta-1,3-galactosyltransferases:
 JOURNAL beta3gal-T5 and beta3gal-T6
 REFERENCE Unpublished
 AUTHORS Amado, M.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAY-1998) Amado M., Department of Oral Diagnostics,

FEATURES
 source Location/Qualifiers
 1. 933
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1. 933
 /gene="beta3gal-T6"
 1. 933
 /gene="beta3gal-T6"
 /codon_start=1
 /product="beta-1,3-galactosyltransferase"
 /protein_id="CAB91547.1"
 /db_xref="GI:7799923"
 /translation="MAPPKRMVLYICLLVLGALCLYMSNLPNPFKEQSFVYKKDGNF
 LKLPDTCROTPDFLLVLTSSSHKQLAERMAIROTGWKERTVKQKLTFFLLGTSS
 AAEKVDQESORHGDIIQKDFLDVYNLTMTMGIEVHRECPQAAEVMKTDSDMF
 INVYLLLELLKNNRTTRFTGKLNNEPFIROFFSKWVSKSEVPWDRYPFPCSGTG
 YVFGSDASOVYVNSVYKIKLEDVFGLCLERLNIRLEELHLSQTPFPFGLRFSVC
 LFRRIVACHFIKPRITLDYWOALENSRGEDCPV"
 BASE COUNT 229 a 234 c 240 g 229 t 1 others
 ORIGIN
 Query Match 38.9%; Score 292; DB 9; Length 933;
 Best Local Similarity 99.5%; Pred. No. 3.3e-166;
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 311 ACAGTCTAATCCTTTCAAGAACAGTCTCTTTGTTTACAGAAAGACGGAACTTCCTTA 370
 Db 77 ACAGTCTAATCCTTTCAAGAACAGTCTCTTTGTTTACAGAAAGACGGAACTTCCTTA 136
 QY 371 AGTCCAGATACAGACTGCAGGACAGACCTCCTCTCTCTCTCTCTCTCTCTCTCTCT 430
 Db 137 AGTCCAGATACAGACTGCAGGACAGACCTCCTCTCTCTCTCTCTCTCTCTCTCTCT 196
 QY 431 CCCACAAACAGTTGGCTGAGCGCATCGGCATCCCGCAGACGCTGGGGAAGAGAGGATGG 490
 Db 197 CCCACAAACAGTTGGCTGAGCGCATCGGCATCCCGCAGACGCTGGGGAAGAGAGGACGG 256
 QY 491 TGAAGGGAAGCAGCTGAAGACATTTCTCTCTCTGGGACACACAGCAGCTGCACGGGAAA 550
 Db 257 TGAAGGGAAGCAGCTGAAGACATTTCTCTCTCTGGGACACACAGCAGCTGCACGGGAAA 316
 QY 551 CGAAGAGCTGACACAGAGAGCCAGCAGCAGCGGACATTCAGAAAGATTTCCTAG 610
 Db 317 CAAGAAGCTGACACAGAGAGCCAGCAGCAGCGGACATTCAGAAAGATTTCCTAG 376
 QY 611 ACCTCTATTACAATCTGACCTGAAGACCATGATGGGCATAGATGGTCCATCGCTTTT 670
 Db 377 ACCTCTATTACAATCTGACCTGAAGACCATGATGGGCATAGATGGTCCATCGCTTTT 436
 QY 671 GTCTCAGCGCGGCTTGTGATGAAGAACAGACTC 704
 Db 437 GTCTCAGCGCGGCTTGTGATGAAGAACAGACTC 470

RESULT 11
 AB041415
 LOCUS Pan paniscus beta1-3-GalT 5 gene for UDP-Gal:GlcNAc
 DEFINITION
 ACCESSION AB041415
 VERSION AB041415.1 GI:7593024
 KEYWORDS
 SOURCE Pan paniscus (isolate:bonobo-05) DNA.
 ORGANISM Pan paniscus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE
 AUTHORS 1 (bases 1 to 1570)
 TITLE Silver Project
 JOURNAL Published Only in Database (2000)
 REFERENCE 2 (bases 1 to 1570)
 AUTHORS Liu, Y. and Saitou, N.

TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL: http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789)

FEATURES Location/Qualifiers
source
1..1570
/organism="Pan paniscus"
/isolate="bonobo-05"
/db_xref="taxon:9597"
/note="human sequence used for primer design based on Acc# AB020337"
<1..20
/number=3
21..667
/number=3
668..1570
/gene="betal.3-Galt 5"
668..>1570
/gene="betal.3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BAA94500.1"
/db_xref="GI:7593025"
/translation="MAPPKRLMYVCLLVGALCLVFSMYSLNLFKEQSFVYKDGNE LKLPDTCROTPEFLVLLVTSSHROLAERMAIQTWGKERTVKGQKLFLLGTTSS AAEKVEDQESQRHGLIQKDFLDGYNLTLMGIEWHRCFOAAFMVKTDSDF INVDYLLKLNRTTRFTGKLNERPPOPFKSFVSKSEYPMWDRYPPFCSGTG YVFGSDVASQVYNSVSEYPIKLEDFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC RFRIRVACHEIKPRTLDDYWQALEN"
668..>1570
/gene="betal.3-Galt 5"
/number=4
/name="4"
408 a 372 c 392 g 398 t

BASE COUNT 408 a 372 c 392 g 398 t

ORIGIN

Query Match 32.1%; Score 241; DB 9; Length 1570;
Best Local Similarity 99.2%; Pred. No. 5e-135;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 358 GGGAACTTCCTTAGCTCCGAGATACAGACTGCAGGAGACACTCCCTCCCTCGTCCTG 417
|||||
DB 791 GGGAACTTCCTTAGCTCCGAGATACAGACTGCAGGAGACACTCCCTCCCTCGTCCTG 850
|||||
QY 418 CTGGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGG 477
|||||
DB 851 CTGGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGG 910
|||||
QY 478 AAAGAGAGATGGTGAAGGGAACAGCTGAAGACATTCCTTCCTCGTGGGACACACAGC 537
|||||
DB 911 AAAGAGAGAGCGTGAAGGGAACAGCTGAAGACATTCCTTCCTCGTGGGACACACAGC 970
|||||
QY 538 AGTGCAGCGGAACGAAGAGGTGGACAGGAGCGGAGCGGAGCGGAGCATATCCAG 597
|||||
DB 971 AGTGCAGCGGAACGAAGAGGTGGACAGGAGCGGAGCGGAGCGGAGCATATCCAG 1030
|||||
QY 598 AAGGATTCCTAGACGTCTATTACACTCTGACCTGAAGACCATGATGGGCATAGATGG 657
|||||
DB 1031 AAGGATTCCTAGACGTCTATTACACTCTGACCTGAAGACCATGATGGGCATAGATGG 1090
|||||
QY 658 GTCCATCGCTTTTCTCTAGCGGGCGTTTGTGATGAAACACAGACTCAGACATGTTTCATC 717
|||||
DB 1091 GTCCATCGCTTTTCTCTAGCGGGCGTTTGTGATGAAACACAGACTCAGACATGTTTCATC 1150
|||||
QY 718 AATGTTGACTATCTGACTGAACTGCTTCTGAAGA 751
|||||
DB 1151 AATGTTGACTATCTGACTGAACTGCTTCTGAAGA 1184
|||||

RESULT 12
AB041414
LOCUS AB041414 1566 bp DNA linear PRI 13-APR-2000

DEFINITION Pan troglodytes betal,3-Galt 5 gene for UDP-Gal:GlcNAc betal,3-galactosyltransferase 5, partial cds.

ACCESSION AB041414
VERSION AB041414.1 GI:7593022
KEYWORDS
SOURCE Pan troglodytes (isolate:chimp-202) DNA.
ORGANISM Pan troglodytes
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1566)
AUTHORS Liu, Y. and Saitou, N.
TITLE Silver Project
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1566)
AUTHORS Liu, Y. and Saitou, N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL: http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789)

FEATURES Location/Qualifiers
source
1..1566
/organism="Pan troglodytes"
/isolate="chimp-202"
/db_xref="taxon:9598"
/note="human sequence used for primer design based on Acc# AB020337"
<1..28
/number=3
29..675
/number=3
675..1566
/gene="betal.3-Galt 5"
675..>1566
/gene="betal.3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal,3-galactosyltransferase 5"
/protein_id="BAA94499.1"
/db_xref="GI:7593023"
/translation="MAPPKRLMYVCLLVGALCLVFSMYSLNLFKEQSFVYKDGNE LKLPDTCROTPEFLVLLVTSSHROLAERMAIQTWGKERTVKGQKLFLLGTTSS AAEKVEDQESQRHGLIQKDFLDGYNLTLMGIEWHRCFOAAFMVKTDSDF INVDYLLKLNRTTRFTGKLNERPPOPFKSFVSKSEYPMWDRYPPFCSGTG YVFGSDVASQVYNSVSEYPIKLEDFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC RFRIRVACHEIKPRTLDDYWQ"
675..>1566
/gene="betal.3-Galt 5"
/number=4
404 a 374 c 394 g 393 t 1 others

BASE COUNT 404 a 374 c 394 g 393 t 1 others

ORIGIN

Query Match 23.4%; Score 176; DB 9; Length 1566;
Best Local Similarity 98.6%; Pred. No. 2.7e-95;
Matches 626; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 117 CAGGAGGTTCTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCTCCTAGCAT 176
|||||
DB 557 CAGGAGGTTCTAGAGTTTCCAAACACGGGTCTCTCTCCACCTCAGCTCCTAGCAT 616
|||||
QY 177 AAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGATTTTTCCTTCAGAT 236
|||||
DB 617 AAAACTAGACACATCCTCATGCTTTTGAGGTCTAATCATTTGATTTTTCCTTCAGAT 676
|||||
QY 237 GGCCTTCCCGAAGATGAGATTGATGATATTTGCGTCTCTGGTCTGGGGGCTCTTGT 296
|||||
DB 677 GGCCTTCCCGAAGATGAGATTGATGATATTTGCGTCTCTGGTCTGGGGGCTCTTGT 736
|||||
QY 297 GTATTTAGCATGTACAGTCTAATCTCTCAAGACAGTCTCTTTTACAGAAAGA 356
|||||
DB 737 GTATTTAGCATGTACAGTCTAATCTCTCAAGACAGTCTCTTTTACAGAAAGA 796
|||||
QY 357 CGGGAATTCCTTAAGCTCCAGATACAGACTGCAGGACACACCTCCCTTCCTCGTCT 416
|||||

```

|||||
Db 797 CGGAACTTCCTTAAGCTCCAGATACAGACTGCAGGACACACCTCCCTCTCGTCT 856
|||||
Qy 417 GCTGGTGACCTATCCCAACACAGTGTGCTGAGCGCATGCCACCGAGAGCTGGG 476
|||||
Db 857 GCTGGTGACCTATCCCAACACAGTGTGCTGAGCGCATGCCACCGAGAGCTGGG 916
|||||
Qy 477 GAAAGAGAGGATGGTGAAGGGAAGCAGCTGAAGACATCTTCCTCCCTGGGACACAC 536
|||||
Db 917 GAAAGAGAGGACGCTGAAGGGAAGCAGCTGAAGACATCTTCCTCCCTGGGACACAC 976
|||||
Qy 537 CAGTGCAGCGGAACGAAGAGAGTGAACAGGAGAGCGACGACGCGGAGCATATCCA 596
|||||
Db 977 CAGTGCAGCGGAACGAAGAGAGTGAACAGGAGAGCGACGACGCGGAGCATATCCA 1036
|||||
Qy 597 GAAGGATTTCTAGAGCTCTATTACAACTCTGACCTCTGAAGACCATGATGGGCATAGA 656
|||||
Db 1037 GAAGGATTTCTAGAGCTCTATTACAACTCTGACCTCTGAAGACCATGATGGGCATAGA 1096
|||||
Qy 657 GGTCCATCGCTTTTGCTCAGCGCGCTTTGTGATGAACAGACTCAGACATCTTCAT 716
|||||
Db 1097 GGTCCATCGCTTTTGCTCAGCGCGCTTTGTGATGAACAGACTCAGACATCTTCAT 1156
|||||
Qy 717 CAATGTTGACTATCTGACTGAATGCTTCTGAAGA 751
|||||
Db 1157 CAATGTTGACTATCTGACTGAATGCTTCTGAAGA 1191
|||||

RESULT 13
AB041417 1579 bp DNA linear PRI 13-APR-2000
LOCUS
DEFINITION
Pongo pygmaeus betal.3-Galt 5 gene for UDP-Gal:GlcNAc
betal.3-galactosyltransferase 5, partial cds.
ACCESSION
AB041417
VERSION
AB041417.1 GI:7593028
KEYWORDS
Pongo pygmaeus (isolate:oran-Pol7) DNA.
ORGANISM
Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
1 (bases 1 to 1579)
Liu, Y. and Saitou, N.
Silver Project
Published Only in DataBase (2000)
2 (bases 1 to 1579)
Liu, Y. and Saitou, N.
Direct Submission
Submitted (11-APR-2000) Naruya Saitou, National Institute of
Genetics, Laboratory of Evolutionary Genetics, 1111 Yata, Mishima,
Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp).
URL: http://sayer.lab.nig.ac.jp/~silver/, Tel: 81-559-81-6790,
Fax: 81-559-81-6789)
FEATURES
Location/Qualifiers
1..1579
/organism="Pongo pygmaeus"
/isolate="oran-Pol7"
/db_xref="taxon:9600"
/notes="human sequence used for primer design based on Acc#
AB020337"
<1..26
/number=3
27..673
/number=3
674..1579
/gene="betal.3-Galt 5"
674..>1579
/gene="betal.3-Galt 5"
/codon_start=1
/product="UDP-Gal:GlcNAc betal.3-galactosyltransferase 5"
/protein_id="BAA94502.1"
/db_xref="GI:7593029"
/translation="MAFPRMLIYICLLVLGALCLYFSMYNLNPFKEOSFYVKKEDGN
FLKLPDTCRQPFPLVLLVTSRHLQALRMAIRQTWGRKTVKQKLTFFLLGTTIS

```

```

SAAMKEVDQESQRHGDIQKDELVDVYVNLTKTMGMENWHREFCPOAAAFVMTDSDM
FINVDYLTLLKKNRTTFFTGFLKLNELPIRQPFKMFVSKSETPWDRYPPFGSGT
AYFSGDVASQVYNVSEVPYIKLEDVFGVGLCLERLNIIRLELHSDPTFFPEGLRFSV
CRFRRIVACHFIKPRTLDDYWALEN"
674..>1579
/gene="betal.3-Galt 5"
/number=4 421 a 371 c 383 g 404 t

Query Match 18.4%; Score 138; DB 9; Length 1579;
Best Local Similarity 99.5%; Pred. No. 4.5e-72;
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 125 TTCTAGAGTTTCCAAACACGGGTCTCTCCACACCTCAGCCTCCTAGCATAAACTAG 184
|||||
Db 564 TTCTAGAGTTTCCAAACACGGGTCTCTCCACACCTCAGCCTCCTAGCATAAACTAG 623
|||||
Qy 185 ACACATCCTCATGCTTTTGAGGTCTAATCATTTGTTTCTCCTTTTCAGATGCTTTCC 244
|||||
Db 624 ACACATCCTCATGCTTTTGAGGTCTAATCATTTGTTTCTCCTTTTCAGATGCTTTCC 683
|||||
Qy 245 CGAAGATGAGATGATGATATTTTCCCTTCCTGTTCTGGGGCTCTTTGTTGTTATTTA 304
|||||
Db 684 CGAAGATGAGATGATGATATTTTCCCTTCCTGTTCTGGGGCTCTTTGTTGTTATTTA 743
|||||
Qy 305 GCATGTACA 313
|||||
Db 744 GCATGTACA 752

RESULT 14
E38439/c Novel polypeptide.
LOCUS
DEFINITION
E38439
ACCESSION
E38439.1 GI:18627013
VERSION
JP 2000245464-A/21.
KEYWORDS
synthetic construct.
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE
1 (bases 1 to 31)
Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
AUTHORS
Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
TITLE
Novel polypeptide
JOURNAL
Patent: JP 2000245464-A 21 12-SEP-2000;
KYOWA HAKKO KOGYO CO LTD
COMMENT
OS Artificial Sequence
PN JP 2000245464-A/21
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
C12P21/02,
PC C12P21/08, C12Q1/68, G01N33/53, (C12N1/21, C12R1/185), (C12N5/10,
C12R1/91),
PC (C12P21/02, C12R1/185), (C12P21/02, C12R1/91), C12N15/00, C12N5/00,
PC (C12N5/00, C12R1/91)
CC
Key Location/Qualifiers
FH source 1..31
FT /organism="Artificial Sequence"
FEATURES
Location/Qualifiers
1..31
source
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT
10 a 5 c 7 g 9 t
ORIGIN

Query Match 3.3%; Score 25; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 305 GCATGACAGTCTAAATCCTTTCAA 329
Db 31 GCATGACAGTCTAAATCCTTTCAA 7

RESULT 15
E38440/c
LOCUS E38440 31 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel polypeptide.
ACCESSION E38440
VERSION E38440.1 GI:18627014
KEYWORDS JP 2000245464-A/22.
SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequences.
1 (bases 1 to 31)
REFERENCE
Narimatsu,H., Isshiki,S., Togayauchi,A. and Sasaki,K.
AUTHORS Novel polypeptide
TITLE Patent: JP 2000245464-A 22 12-SEP-2000;
JOURNAL KYOWA HAKKO KOGYO CO LTD
COMMENT OS Artificial Sequence
PN JP 2000245464-A/22
PD 12-SEP-2000
PF 25-FEB-1999 JP 1999047571
PR
PI HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
KATSUTOSHI SASAKI
PC C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12P19/00, PC
C12P21/02,
PC C12P21/08,C12Q1/68,G01N33/53//(C12N1/21,C12R1:185), (C12N5/10,
PC C12R1:91),
PC (C12P21/02,C12R1:185), (C12P21/02,C12R1:91),C12N15/00,C12N5/00,
PC (C12N5/00,C12R1:91)
CC
FH
FT
FT source
FEATURES
Location/Qualifiers
1..31
1..31 /organism="Artificial Sequence"
/db_xref="taxon:32630"
BASE COUNT 4 a 8 c 8 g 11 t
ORIGIN

Query Match 3.2%, Score 24; DB 6; Length 31;
Best Local Similarity 100.0%; Pred.No. 0.016;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 376 CCAGATACAGACTGCAGGCAGACA 399
Db 31 CCAGATACAGACTGCAGGCAGACA 8
```

Search completed: April 12, 2003, 08:09:49
Job time : 1603.57 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: April 11, 2003, 22:38:14 ; Search time 123.291 Seconds
(without alignments)
13717.504 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750
Perfect score: 751
Sequence: 1 ggaccaaagccagaatctc.....gactgaactgtttctgaaga 751

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	100.0	10562	21 AAA93876	Human beta3Gal-T5
2	517	68.8	2775	21 AAA93875	Human beta-1,3 gal
3	442	58.9	1011	21 AAA27959	Human beta3Gal-T5
C 4	25	3.3	31	21 AAA93894	Beta-Gal-T5 specif
C 5	24	3.2	31	21 AAA93895	Beta-Gal-T5 specif
6	23	3.1	23	21 AAA93877	Beta-1, 3 galactos
7	23	3.1	23	21 AAA93905	Forward PCR primer
8	21	2.8	21	21 AAA27952	PCR primer EBER 13
9	21	2.8	21	21 AAA27954	PCR primer EBER 13

10	21	2.8	21	21 AAA27957	PCR primer EBER 13
C 11	20	2.7	318	22 AA182687	Human polynucleoti
C 12	20	2.7	405	24 ABL86566	Human ovarian canc
C 13	20	2.7	411	22 AAK69094	Human immune/haema
C 14	20	2.7	489	22 AAK70504	Human immune/haema
C 15	20	2.7	572	22 AAK69488	Human immune/haema
C 16	20	2.7	1155	22 AAK81186	Human immune/haema
C 17	20	2.7	1155	22 AAK84745	Human immune/haema
C 18	20	2.7	2170	22 AAD02698	Human glycosyl sul
C 19	20	2.7	2928	24 ABK34615	Human cDNA for nov
C 20	20	2.7	2980	22 ABA17920	Human nervous syst
C 21	20	2.7	2988	21 AAC76156	Human OREF ORF1711
C 22	20	2.7	3784	22 AAL07327	Human reproductive
C 23	20	2.7	3784	22 AAL07330	Human reproductive
C 24	20	2.7	5377	22 AAL08680	Human breast or ov
C 25	20	2.7	5377	22 AAL08686	Human breast or ov
C 26	20	2.7	7389	22 AAS27788	DNA encoding novel
C 27	20	2.7	10953	22 AAS29204	Genomic sequence #
C 28	20	2.7	11293	22 ABA17918	Human nervous syst
C 29	20	2.7	13337	22 ABA18582	Human nervous syst
C 30	20	2.7	13337	22 AAS29203	Genomic sequence #
C 31	20	2.7	13337	23 ABK42742	Genomic sequence #
C 32	20	2.7	14049	22 ABA15812	Human nervous syst
C 33	20	2.7	15000	24 ABL63905	Breast cancer rela
C 34	20	2.7	16920	22 AAK69093	Human immune/haema
C 35	20	2.7	22107	22 ABA17921	Human nervous syst
C 36	20	2.7	22107	22 AAK79542	Human immune/haema
C 37	20	2.7	22109	22 ABA17922	Human nervous syst
C 38	20	2.7	22109	22 AAK79543	Human immune/haema
C 39	20	2.7	22111	22 ABA17919	Human nervous syst
C 40	20	2.7	22111	22 AAK79540	Human immune/haema
C 41	20	2.7	22916	22 AAK65305	Human immune/haema
C 42	20	2.7	27869	22 ABA19635	Human nervous syst
C 43	20	2.7	27869	22 AAK66517	Human immune/haema
C 44	20	2.7	32134	22 ABA15354	Human nervous syst
C 45	20	2.7	32134	22 ABA15813	Human nervous syst

ALIGNMENTS

RESULT 1
AAA93876
ID AAA93876 standard; DNA; 10562 BP.
XX
AC AAA93876;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta3Gal-T5 encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; beta3Gal-T5; ds.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PR (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer

XX PS Claim 31; Page 103-111; 123pp; Japanese.

CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents a Beta3gal-T5
CC encoding DNA sequence.

XX SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 100.0%; Score 751; DB 21; Length 10562;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGCG 60
DB 8000 GGACCAATGCCAGAACTCTGAGCTTTTATCTTACACCATGAAGTGACAGATGCTGCG 8059

QY 61 AGATCTTAGACCTTTGCTTAAGTCTTAACACACAGACACCGGACTTCTGTATGACG 120
DB 8060 AGATCTTAGACCTTTGCTTAAGTCTTAACACACAGACACCGGACTTCTGTATGACG 8119

QY 121 GAGGTTCTAGAGTTTCCAAACAGCGGTCTCTCTCCACCTCAGGCTCCTAGCAATAAAA 180
DB 8120 GAGGTTCTAGAGTTTCCAAACAGCGGTCTCTCTCCACCTCAGGCTCCTAGCAATAAAA 8179

QY 181 CTAGACATCTCATGCTTTGAGGCTTAATCATGGATTTTGTTCCTTTCAGATGGCT 240
DB 8180 CTAGACATCTCATGCTTTGAGGCTTAATCATGGATTTTGTTCCTTTCAGATGGCT 8239

QY 241 TTCGCGAAGATGAGATTCATGTATTTGGCTTCTGTGTTCTGGGGGCTCTTTTGTAT 300
DB 8240 TTCGCGAAGATGAGATTCATGTATTTGGCTTCTGTGTTCTGGGGGCTCTTTTGTAT 8299

QY 301 TTTAGCATGTACGCTAAATCTTTTCAAAGACAGTCCTTTGTTTACAGAAGACGGG 360
DB 8300 TTTAGCATGTACGCTAAATCTTTTCAAAGACAGTCCTTTGTTTACAGAAGACGGG 8359

QY 361 AACTTCTTAAAGTCCAGATACAGCTGCAGGACAGACCTCCTTCTCGTCCTGCTG 420
DB 8360 AACTTCTTAAAGTCCAGATACAGCTGCAGGACAGACCTCCTTCTCGTCCTGCTG 8419

QY 421 GTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGGGAAA 480
DB 8420 GTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGGGGGAAA 8479

QY 481 GAGAGGATGTTGAAGGAAAGCAGCTCAAGACATTTCTTCTCTGGGGACACACGACGT 540
DB 8480 GAGAGGATGTTGAAGGAAAGCAGCTCAAGACATTTCTTCTCTGGGGACACACGACGT 8539

QY 541 GCAGCGGAAGAAAGAGTTGGACGAGAGAGCCAGCGACGCGGACATATTCCAGAAG 600
DB 8540 GCAGCGGAAGAAAGAGTTGGACGAGAGAGCCAGCGACGCGGACATATTCCAGAAG 8599

QY 601 GATTTCCTTAGAGCTCTATTACATCTCACCTCGAAGACCATGATGGGCATAGATGGTTC 660
DB 8600 GATTTCCTTAGAGCTCTATTACATCTCACCTCGAAGACCATGATGGGCATAGATGGTTC 8659

QY 661 CATGCTTTTCTCTCAGGGGCGGTTTGTGATGAAACAGACACTACAGATGTTTCATCAAT 720
DB 8660 CATGCTTTTCTCTCAGGGGCGGTTTGTGATGAAACAGACACTACAGATGTTTCATCAAT 8719

QY 721 GTTGACTATCTGACTGAAGTCTTCTGAAGA 751

DB 8720 GTTGACTATCTGACTGAAGTCTTCTGAAGA 8750

RESULT 2
AAA93875
ID AAA93875 standard; DNA; 2775 BP.
XX AAA93875;
AC AAA93875;
XX 15-JAN-2001 (first entry)
XX Human beta-1,3 galactose transferase encoding DNA.
XX Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; ds.
XX Homo sapiens.
XX WO2000050608-A1.
XX 31-AUG-2000.
XX 24-FEB-2000; 2000WO-JP01070.
XX 25-FEB-1999; 99JP-0047571.
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Narimatsu H, Ishiki S, Togayachi A, Sasaki K;
PI WPI; 2000-549409/50.
DR P-PSDB; AAB93875.
XX Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer -
XX Claim 5; Page 99-102; 123pp; Japanese.
XX This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents Beta-1,3 galactose
CC transferase encoding DNA.

XX SQ Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 other;

Query Match 68.8%; Score 517; DB 21; Length 2775;
Best Local Similarity 100.0%; Pred. No. 1.6e-257;
Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 ATGGCTTTCCCGAAGATGAGATTCATGTATATTTCCTTCTGTTCTGGGGGCTCTTTGT 294
DB 402 ATGGCTTTCCCGAAGATGAGATTCATGTATATTTCCTTCTGTTCTGGGGGCTCTTTGT 461

QY 295 TTGATTTTACATGTACAGTCTAAATCCTTTCAGAACAGTCCCTTTGTTTACAGAAA 354
DB 462 TTGATTTTACATGTACAGTCTAAATCCTTTCAGAACAGTCCCTTTGTTTACAGAAA 521

QY 355 GACGGGAACCTTCCCTTAAGCTCCAGATACAGCTGCAGGACAGACCTCCCTCCCTCGTC 414
DB 522 GACGGGAACCTTCCCTTAAGCTCCAGATACAGCTGCAGGACAGACCTCCCTCCCTCGTC 591

QY 415 CTGCTGTGACCTCATCCCAACACAGTTGGCTGAGCGCATGGCCATCCGGCAGAGCTGG 474

Db 582 CTGCTGGTGACCTCATCCCAACACAGTTGGCTGAGGGCATGCCATCCGGCAGACGTGG 641

Qy 475 GGGAAAGAGAGGATGGTGAAGGAAAGACAGCTGAAGACATTTCTCTCTCTGGGACACC 534

Db 642 GGGAAAGAGAGGATGGTGAAGGAAAGACAGCTGAAGACATTTCTCTCTCTGGGACACC 701

Qy 535 AGCAGTCAGCGGAAACGAAGAGGTGGACAGGAGCCAGGACACGGGACATTATC 594

Db 702 AGCAGTCAGCGGAAACGAAGAGGTGGACAGGAGCCAGGACACGGGACATTATC 761

Qy 595 CAGAAGGATTTCTAGACGCTTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 654

Db 762 CAGAAGGATTTCTAGACGCTTATTACAATCTGACCCCTGAAGACCATGATGGGCATAGAA 821

Qy 655 TGGCTCCATCGCTTTCTCTCTAGCGGGCTTTGTGATGAACACACATCAGACATGTTTC 714

Db 822 TGGCTCCATCGCTTTCTCTCTAGCGGGCTTTGTGATGAACACACATCAGACATGTTTC 881

Qy 715 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGA 751

Db 882 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGA 918

RESULT 3

AAA27959

ID AAA27959 standard; DNA; 1011 BP.

XX AAA27959;

AC

XX

XX

DT 15-AUG-2000 (first entry)

XX

DE Human beta3Gal-T5 gene sequence.

XX

KW UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;

KW Beta3Gal-T5; ss; human; Chromosome 21q22.3; galactosylation;

KW beta1,3-galactosyl glycosylated saccharide production; glycopeptide;

KW glycoprotein.

XX

OS Homo sapiens.

XX

XX

Key Location/Qualifiers

FT CDS 79..1011

FT /tag= a

FT /product= "Beta3Gal-T5"

FT primer_bind complement (79..98)

FT /tag= b

FT primer_bind complement (150..170)

FT /tag= c

FT primer_bind 991..1011

FT /tag= d

XX

PN WC200002958-A1.

XX

XX

XX 25-MAY-2000.

XX

XX 11-NOV-1999; 99NO-US26807.

XX

XX 13-NOV-1998; 98DK-0001483.

XX

XX (CLAU/) CLAUSEN H.

XX

XX Clausen H, Anado M;

XX

XX WPI; 2000-399728/34.

DR P-PSDB; AAY94641.

XX

XX Novel nucleic acid sequence encoding human

PT UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase

PT useful for obtaining beta1,3-galactosyl glycosylated saccharides and

PT glycopeptides or glycoproteins

XX

PS Claim 7; Fig 1; 74pp; English.

XX

CC The present invention relates to a nucleic acid sequence encoding

CC UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase

CC (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at

CC carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence

CC represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is

CC located on human chromosome 21q22.3. Beta3Gal-T5 is a type II

CC transmembrane glycoprotein. The invention also relates to the beta3Gal-T5

CC protein sequence, a nucleic acid vector comprising the beta3Gal-T5

CC nucleotide sequence, a host cell comprising the vector, and a method for

CC the production of the beta3Gal-T5 protein from the host cells. The

CC methods of the invention can be used for recombinant production of

CC beta3Gal-T5 for use as a catalyst and for recombinant production of

CC peptides or proteins with appropriate galactosylation. The beta3Gal-T5

CC protein can be used to obtain beta1,3-galactosyl glycosylated

CC saccharides, glycopeptides or glycoproteins.

XX

SQ Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 other;

Query Match 58.9%; Score 442; DB 21; Length 1011;

Best Local Similarity 99.5%; Pred. No. 1.1e-218;

Matches 592; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 157 CCACCTCAGCCTCCTAGCATAAAACACTAGACACATCCTCATCTCTTTGAGTCTAATCATT 216

Db 1 CCACCTCAGCCTCCTAGCATAAAACACTAGACACATCCTCATCTCTTTGAGTCTAATCATT 60

Qy 217 GGATTTGTTTCCCTTCAGATGGCTTTCCCGAAGATGATGATGATATATTGCTTTCTG 276

Db 61 GGATTTGTTTCCCTTCAGATGGCTTTCCCGAAGATGATGATGATATATTGCTTTCTG 120

Qy 277 GTTCTGGGGCTCTTTGTTTGTATTTTAGCATGTACAGTCTAAATCCTTTTCAAGACACAG 336

Db 121 GTTCTGGGGCTCTTTGTTTGTATTTTAGCATGTACAGTCTAAATCCTTTTCAAGACACAG 180

Qy 337 TCCTTTGTTTACAAGAAGACGGGAACCTTCTTAAGTCCCGATACAGACTCGACGCGAG 396

Db 181 TCCTTTGTTTACAAGAAGACGGGAACCTTCTTAAGTCCCGATACAGACTCGACGCGAG 240

Qy 397 ACACCTCCCTTCCTCTGCTGCTGGTGACCTCATCCACAAACAGTTGGCTGAGCGCATG 456

Db 241 ACACCTCCCTTCCTCTGCTGCTGGTGACCTCATCCACAAACAGTTGGCTGAGCGCATG 300

Qy 457 GCCATCCGGCAGACGTGGGGAAAGAGAGAGTGGTGAAGGAAAGAGAGCTGAAGACATTC 516

Db 301 GCCATCCGGCAGACGTGGGGAAAGAGAGAGCGGTGAAGGAAAGAGAGCTGAAGACATTC 360

Qy 517 TTCCTCTCTGGGACCCACGACGTGCAGCGGAAACGAAAGAGGTGGACACGAGAGCCAG 576

Db 361 TTCCTCTCTGGGACCCACGACGTGCAGCGGAAACGAAAGAGGTGGACACGAGAGCCAG 420

Qy 577 CGACACGGGACATTTATCCAGAGGATTTCTTAGACGCTCTATTACATCTGACCCCTGAAG 636

Db 421 CGACACGGGACATTTATCCAGAGGATTTCTTAGACGCTCTATTACATCTGACCCCTGAAG 480

Qy 637 ACCATGATGGGCATAGAAATGGGTCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAAA 696

Db 481 ACCATGATGGGCATAGAAATGGGTCATCGCTTTTGTCTCAGCGCGGCTTTGTGATGAAA 540

Qy 697 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTTCTCTGAAGA 751

Db 541 ACAGACTCAGACATGTTTCATCAATGTTGACTATCTGACTGAATGCTTCTCTGAAGA 595

RESULT 4

AAA93894/C

ID AAA93894 standard; DNA; 31 BP.

XX AAA93894;

AC

XX

DT 15-JAN-2001 (first entry)

XX

DE Beta-Gal-T5 specific PCR primer SEQ ID 22.

XX

KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
 KW digestive system; PCR primer; ss.
 OS Synthetic.
 XX WO200050608-A1.
 XX 31-AUG-2000.
 XX 24-FEB-2000; 2000WO-JP01070.
 XX 25-FEB-1999; 99JP-0047571.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
 XX WPI; 2000-549409/50.
 XX Beta-1,3 galactose transferase and DNA encoding it, useful for
 PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
 PT digestive system cancer -
 XX Example 5; Page 118; 123pp; Japanese.
 XX This invention relates to a polypeptide (I) with beta-1,3 galactose
 CC transferase activity, or variants of (I) comprising amino acid additions,
 CC deletions and/or substitutions. Included in the invention is DNA encoding
 CC all or part of (I); expression vectors containing the DNA, host cells
 CC transformed by the vectors; a method for the preparation of the
 CC polypeptide by culture of the transformants or by expression in the milk
 CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
 CC galactose transferase protein transfers galactose by beta-1,3 bonding to
 CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
 CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
 CC DNA encoding it are useful for the treatment and diagnosis of cancer of
 CC the digestive system. The present sequence represents a Beta-gal-T5
 CC specific PCR primer used in the invention.
 XX SQ Sequence 31 BP; 10 A; 5 C; 7 G; 9 T; 0 other;
 Query Match 3.3%; Score 25; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 305 GCATGTACAGTCTAAATCCCTTCAA 329
 Db 31 GCATGTACAGTCTAAATCCCTTCAA 7
 RESULT 5
 AAA93895/C
 ID AAA93895 standard; DNA; 31 BP.
 XX
 AC AAA93895;
 XX 15-JAN-2001 (first entry)
 DT Beta-Gal-T5 specific PCR primer SEQ ID 23.
 DE Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
 KW digestive system; PCR primer; ss.
 XX Synthetic.
 OS WO200050608-A1.
 XX 31-AUG-2000.
 XX 24-FEB-2000; 2000WO-JP01070.
 XX 25-FEB-1999; 99JP-0047571.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
 XX WPI; 2000-549409/50.
 XX Beta-1,3 galactose transferase and DNA encoding it, useful for
 PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
 PT digestive system cancer -
 XX Claim 28; Page 117; 123pp; Japanese.
 XX This invention relates to a polypeptide (I) with beta-1,3 galactose

PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
 XX WPI; 2000-549409/50.
 XX Beta-1,3 galactose transferase and DNA encoding it, useful for
 PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
 PT digestive system cancer -
 XX Example 5; Page 118; 123pp; Japanese.
 XX This invention relates to a polypeptide (I) with beta-1,3 galactose
 CC transferase activity, or variants of (I) comprising amino acid additions,
 CC deletions and/or substitutions. Included in the invention is DNA encoding
 CC all or part of (I); expression vectors containing the DNA, host cells
 CC transformed by the vectors; a method for the preparation of the
 CC polypeptide by culture of the transformants or by expression in the milk
 CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
 CC galactose transferase protein transfers galactose by beta-1,3 bonding to
 CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
 CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
 CC DNA encoding it are useful for the treatment and diagnosis of cancer of
 CC the digestive system. The present sequence represents a Beta-Gal-T5
 CC specific PCR primer used in the invention.
 XX SQ Sequence 31 BP; 4 A; 8 C; 8 G; 11 T; 0 other;
 Query Match 3.2%; Score 24; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 0.046;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 376 CCAGATACAGACTGCAGGCAGACA 399
 Db 31 CCAGATACAGACTGCAGGCAGACA 8
 RESULT 6
 AAA93877
 ID AAA93877 standard; DNA; 23 BP.
 XX
 AC AAA93877;
 XX 15-JAN-2001 (first entry)
 DT Beta-1, 3 galactose transferase PCR primer SEQ ID 20.
 DE Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
 KW digestive system; PCR primer; ss.
 XX Synthetic.
 OS WO200050608-A1.
 XX 31-AUG-2000.
 XX 24-FEB-2000; 2000WO-JP01070.
 XX 25-FEB-1999; 99JP-0047571.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
 XX WPI; 2000-549409/50.
 XX Beta-1,3 galactose transferase and DNA encoding it, useful for
 PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
 PT digestive system cancer -
 XX Claim 28; Page 117; 123pp; Japanese.
 XX This invention relates to a polypeptide (I) with beta-1,3 galactose

CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents a Beta-1,3
CC galactose transferase PCR primer.

XX SQ Sequence 23 BP; 8 A; 8 C; 6 G; 1 T; 0 other;

Query Match 3.1%; Score 23; DB 21; Length 23;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 ACCACCAGCAGTCGACGGGAAC 551

|||||

Db 1 ACCACCAGCAGTCGACGGGAAC 23

RESULT 7

AAA93905

ID AAA93905 standard; DNA; 23 BP.

AC AAA93905;

XX

XX

DT 15-JAN-2001 (first entry)

XX

DE Forward PCR primer for Beta3Gal-T5 amplification.

XX

KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;

KW digestive system; PCR primer; ss.

XX

OS Synthetic.

XX

PN WO200050608-A1.

XX

XX

PF 24-FEB-2000; 2000WO-JP01070.

XX

PR 25-FEB-1999; 99JP-0047571.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX

PI Narimatsu H, Ishiki S, Togayachi A, Sasaki K;

XX

DR WPI; 2000-549409/50.

XX

PT Beta-1,3 galactose transferase and DNA encoding it, useful for

PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of

XX

XX

PS Example 1; Page 58; 123pp; Japanese.

XX

CC This invention relates to a polypeptide (I) with beta-1,3 galactose

CC transferase activity, or variants of (I) comprising amino acid additions,

CC deletions and/or substitutions. Included in the invention is DNA encoding

CC all or part of (I); expression vectors containing the DNA, host cells

CC transformed by the vectors; a method for the preparation of the

CC polypeptide by culture of the transformants or by expression in the milk

CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3

CC galactose transferase protein transfers galactose by beta-1,3 bonding to

CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as

CC GlcNAc-beta1-3gal-beta1-4Glc) to give Gal-beta1-3GlcNAc. The protein and

SQ Sequence 23 BP; 8 A; 8 C; 6 G; 1 T; 0 other;

Query Match 3.1%; Score 23; DB 21; Length 23;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 ACCACCAGCAGTCGACGGGAAC 551

|||||

Db 1 ACCACCAGCAGTCGACGGGAAC 23

RESULT 8

AAA27952

ID AAA27952 standard; DNA; 21 BP.

XX

AC AAA27952;

XX

DT 15-AUG-2000 (first entry)

XX

DE PCR primer EBER 1301 used for human Beta3Gal-T5 detection.

XX

KW UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
KW Beta3Gal-T5; PCR primer; ss; human; chromosome 21q22.3; galactosylation;
KW beta1,3-galactosyl glycosylated saccharide production; glycopeptide;
KW glycoprotein.

XX

OS Homo sapiens.

XX

PN WO200029558-A1.

XX

XX

PD 25-MAY-2000.

XX

PE 11-NOV-1999; 99WO-US26807.

XX

PR 13-NOV-1998; 98DK-0001483.

XX

PA (CLAU/) CLAUSEN H.

XX

PI Clausen H, Amado M;

XX

DR WPI; 2000-399728/34.

XX

PT Novel nucleic acid sequence encoding human

PT UDP-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase

PT useful for obtaining beta 1,3-galactosyl glycosylated saccharides and

PT glycopeptides or glycoproteins

XX

PS Example; Page 35; 74pp; English.

XX

CC The present invention relates to a nucleic acid sequence encoding

CC UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase

CC (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at

CC carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence

CC represents a PCR primer used to screen a human foreskin fibroblast pl

CC library for the human beta3Gal-T5 sequence. The beta3Gal-T5 gene is

CC located on human chromosome 21q22.3. Beta3Gal-T5 is a type II

CC transmembrane glycoprotein. The invention also relates to the beta3Gal-T5

CC nucleotide sequence, a nucleic acid vector comprising the beta3Gal-T5

CC the production of the beta3Gal-T5 protein from the host cells. The

CC methods of the invention can be used for recombinant production of

CC beta3Gal-T5 for use as a catalyst and for recombinant production of

CC peptides or proteins with appropriate galactosylation. The beta3Gal-T5

CC protein can be used to obtain beta1,3-galactosyl glycosylated

CC saccharides, glycopeptides or glycoproteins.

XX

SQ Sequence 21 BP; 5 A; 8 C; 2 G; 6 T; 0 other;

Query Match

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 CTTCCTTAAGTCCCGATAC 383

```

Db      1 CTTCTTAAGCTCCAGATAC 21
          |||
RESULT 9
AAA27954
ID      AAA27954 standard; DNA; 21 BP.
XX
XX      AAA27954;
XX
XX      15-AUG-2000 (first entry)
XX
XX      PCR primer EBER 1300sol for amplification of a beta3Gal-T5 fragment.
XX
XX      UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
XX      beta3Gal-T5; PCR primer; ss; human; chromosome 21q22.3; galactosylation;
XX      beta1,3-galactosyl glycosylated saccharide production; glycopeptide;
XX      glycoprotein.
XX
XX      Homo sapiens.
XX
XX      WO200029558-A1.
XX
XX      25-MAY-2000.
XX
XX      11-NOV-1999; 99WO-US26807.
XX
XX      13-NOV-1998; 98DK-0001483.
XX      (CLAU/) CLAUSEN H.
XX      Clausen H, Amado M;
XX
XX      WPI: 2000-399728/34.
XX
XX      Novel nucleic acid sequence encoding human
XX      UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
XX      useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
XX      glycopeptides or glycoproteins -
XX
XX      Example; Page 36; 74pp; English.
XX
XX      The present invention relates to a nucleic acid sequence encoding
XX      UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
XX      (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
XX      carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
XX      represents a PCR primer used to amplify a fragment of the human
XX      beta3Gal-T5 nucleotide sequence. The beta3Gal-T5 gene is located on human
XX      chromosome 21q22.3. Beta3Gal-T5 is a type II transmembrane glycoprotein.
XX      The invention also relates to the beta3Gal-T5 protein sequence, a nucleic
XX      acid vector comprising the beta3Gal-T5 nucleotide sequence, a host cell
XX      comprising the vector, and a method for the production of the beta3Gal-T5
XX      protein from the host cells. The methods of the invention can be used for
XX      recombinant production of beta3Gal-T5 for use as a catalyst and for
XX      galactosylation. The beta3Gal-T5 protein can be used to obtain
XX      beta1,3-galactosyl glycosylated saccharides, glycopeptides or
XX      glycoproteins.
XX
XX      Sequence 21 BP; 6 A; 5 C; 2 G; 8 T; 0 other;
XX
XX      Query Match 2.8%; Score 21; DB 21; Length 21;
XX      Best Local Similarity 100.0%; Pred. No. 1.6;
XX      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      307 ATGTACAGTCTAATCCTTTC 327
          |||
Db      1 ATGTACAGTCTAATCCTTTC 21
          |||
RESULT 10
AAA27957
ID      AAA27957 standard; DNA; 21 BP.
XX
XX      15-AUG-2000 (first entry)
XX
XX      PCR primer EBER 1320 used in polymorphism analysis of beta3Gal-T5 gene.
XX
XX      UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
XX      beta3Gal-T5; PCR primer; ss; human; chromosome 21q22.3; galactosylation;
XX      beta1,3-galactosyl glycosylated saccharide production; glycopeptide;
XX      glycoprotein.
XX
XX      Homo sapiens.
XX
XX      WO200029558-A1.
XX
XX      25-MAY-2000.
XX
XX      11-NOV-1999; 99WO-US26807.
XX
XX      13-NOV-1998; 98DK-0001483.
XX      (CLAU/) CLAUSEN H.
XX      Clausen H, Amado M;
XX
XX      WPI: 2000-399728/34.
XX
XX      Novel nucleic acid sequence encoding human
XX      UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
XX      useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
XX      glycopeptides or glycoproteins -
XX
XX      Example; Page 42; 74pp; English.
XX
XX      The present invention relates to a nucleic acid sequence encoding
XX      UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
XX      (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
XX      carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
XX      represents a PCR primer used in the analysis of DNA polymorphisms in the
XX      beta3Gal-T5 gene. The beta3Gal-T5 gene is located on human chromosome
XX      21q22.3. Beta3Gal-T5 is a type II transmembrane glycoprotein. The
XX      invention also relates to the beta3Gal-T5 protein sequence, a nucleic
XX      acid vector comprising the beta3Gal-T5 nucleotide sequence, a host cell
XX      comprising the vector, and a method for the production of the beta3Gal-T5
XX      protein from the host cells. The methods of the production of the beta3Gal-T5
XX      protein from the host cells. The methods of the invention can be used for
XX      recombinant production of beta3Gal-T5 for use as a catalyst and for
XX      galactosylation. The beta3Gal-T5 protein can be used to obtain
XX      beta1,3-galactosyl glycosylated saccharides, glycopeptides or
XX      glycoproteins.
XX
XX      Sequence 21 BP; 4 A; 5 C; 6 G; 6 T; 0 other;
XX
XX      Query Match 2.8%; Score 21; DB 21; Length 21;
XX      Best Local Similarity 100.0%; Pred. No. 1.6;
XX      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      117 CAGCGAGGTTCTAGAGTTTC 137
          |||
Db      1 CAGCGAGGTTCTAGAGTTTC 21
          |||
RESULT 11
AAI82687/c
ID      AAI82687 standard; cDNA; 318 BP.
XX
XX      AAI82687;
XX
XX      06-NOV-2001 (first entry)
XX
XX      Human polynucleotide SEQ ID NO 2747.
XX

```

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX Homo sapiens.
 OS
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 PF
 XX 28-FEB-2000; 2000US-0515126.
 PR
 XX 18-MAY-2000; 2000US-0577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI: 2001-514838/56.
 XX
 DR P-PSDB; AAO02756.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders.
 PT
 XX
 XX Claim 1; SEQ ID NO 2747; 1399pp + Sequence Listing; English.
 PS
 XX
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 318 BP; 86 A; 51 C; 72 G; 109 T; 0 other;
 Query Match 2.7%; Score 20; DB 22; Length 318;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 153 TCTCCACCTCAGCCTCCTA 172
 Db 28 TCTCCACCTCAGCCTCCTA 9
 RESULT 12
 ABL86656/C
 ID ABL86656 standard; cDNA; 405 BP.
 XX
 XX ABL86656;
 AC
 XX
 XX 17-MAY-2002 (first entry)
 DT
 XX Human ovarian cancer related cDNA clone SEQ ID NO:9634.
 DE
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200192581-A2.
 PN
 XX 06-DEC-2001.
 PD
 XX

PF 29-MAY-2001; 2001WO-US17756.
 XX
 PR 26-MAY-2000; 2000US-207484P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Algate PA, Harlocker SL, Jones R;
 PI
 XX WPI: 2002-122075/16.
 DR
 XX
 XX Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide
 XX
 XX Claim 1; SEQ ID 9634; 489pp; English.
 PS
 XX
 XX The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
 CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
 CC population of (III), or antigen presenting cells that express (II).
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
 CC (S1) can be used for detecting ovarian cancer in a patient's biological
 CC sample preferably serum or ovarian tissue. The method comprises
 CC contacting a biological sample from a patient with (IV), detecting the
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the
 CC patient, where the amount of polynucleotide hybridising to (IV) is
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells
 CC specific for an ovarian tumour protein comprising contacting T cells
 CC with (III) or (II). (III) is useful in design and preparation of
 CC ribosome molecules for inhibiting expression of the tumour polypeptides
 CC and proteins in tumour cells; and to isolate a full length gene from a
 CC suitable library e.g., a tumour cDNA library using well known
 CC techniques.
 XX
 SQ Sequence 405 BP; 114 A; 84 C; 125 G; 82 T; 0 other;
 Query Match 2.7%; Score 20; DB 24; Length 405;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 152 CTCTCCACCTCAGCCTCCT 171
 Db 347 CTCTCCACCTCAGCCTCCT 328
 RESULT 13
 AAK69094/C
 ID AAK69094 standard; DNA; 411 BP.
 XX
 XX AAK69094;
 AC
 XX
 XX 06-NOV-2001 (first entry)
 DT
 XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23906.
 DE
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO200157182-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX
 XX 17-JAN-2001; 2001WO-US01354.
 PF
 XX 31-JAN-2000; 2000US-0179065.
 PR
 XX 04-FEB-2000; 2000US-0180628.
 PR

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPT; 2001-483426/52.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
Disclosure; SEQ ID NO 23906; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 XX Sequence 411 BP; 145 A; 82 C; 84 G; 100 T; 0 other;

Query Match 2.7%; Score 20; DB 22; Length 411;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 TCTCCACCTCAGCTCCTA 172
 ||||||||||||||||
 DB 54 TCTCCACCTCAGCTCCTA 35

RESULT 14

AAK70504/c
 ID AAK70504 standard; DNA; 489 BP.
 XX AAK70504;
 AC
 XX
 XX 06-NOV-2001 (first entry)
 DT
 XX
 XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25316.
 DE
 XX
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200157182-A2.
 PN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 17-JAN-2001; 2001WC-US01354.
 PF
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225477.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226379.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233083.
 PR 14-SEP-2000; 2000US-0233084.
 PR 14-SEP-2000; 2000US-0233085.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-483426/52.
DR
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
XX
XX Disclosure; SEQ ID NO 25316; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 489 BP; 139 A; 99 C; 125 G; 126 T; 0 other;

Query Match 2.7%; Score 20; DB 22; Length 489;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 153 TCTCCACCTCAGCTCCTA 172
|||||
Db 312 TCTCCACCTCAGCTCCTA 293
|||||
RESULT 15
AAK69488/C
ID AAK69488 standard; DNA: 572 BP.
XX AC AAK69488;
XX DT 06-NOV-2001 (first entry)
XX DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:24300.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX WC200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

```
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249607.
PR 17-NOV-2000; 2000US-0249613.
PR 17-NOV-2000; 2000US-0249628.
PR 17-NOV-2000; 2000US-0249629.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM.
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
XX Disclosure; SEQ ID NO 24300; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 572 BP; 162 A; 106 C; 119 G; 185 T; 0 other;
Query Match 2.7%; Score 20; DB 22; Length 572;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 153 TCTCCACCTCAGCTCCTTA 172
Db 528 TCTCCACCTCAGCTCCTTA 509
Search completed: April 12, 2003, 03:54:05
Job time : 133.291 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 02:50:35 ; Search time 824.43 Seconds
(without alignments)
14752.993 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750

Perfect score: 751

Sequence: 1 ggaccaatgccagaatctc.....gactgaactgcttctgaaga 751

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 50.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estmv:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
c 1	165	22.0	338	9 AJ003597	AJ003597
2	23	3.1	420	17 AQ094539	AQ094539 HS_3026_B
3	23	3.1	429	14 H17163	H17163 ym37c02.r1
4	23	3.1	432	12 BG152895	BG152895 nad09h06.
5	23	3.1	636	13 BG965086	BG965086 602829157
6	23	3.1	641	13 BG913391	BG913391 602812247

7	23	3.1	674	17	AQ634393	AQ634393 RPCI-11-4
8	21	2.8	286	10	BB284950	BB284950
9	21	2.8	293	10	AW901629	AW901629 PCO-NN101
10	21	2.8	325	9	AI368142	AI368142 q944907.x
11	21	2.8	497	17	AQ368646	AQ368646 HS_5038.A
12	21	2.8	512	10	AW118817	AW118817 xd97a09.x
13	21	2.8	523	14	BM967143	BM967143 j33a05.y
14	21	2.8	523	14	BM967160	BM967160 j33b11.y
15	21	2.8	537	9	AI281699	AI281699 qt67f11.x
16	21	2.8	546	17	AQ211769	AQ211769 HS_3234.B
17	21	2.8	554	14	BM893762	BM893762 j29a01.x
18	21	2.8	560	14	BM732402	BM732402 K-EST0072
19	21	2.8	566	14	BM967046	BM967046 j29a01.y
20	21	2.8	646	17	AZ519595	AZ519595 RPCI-11-2
21	21	2.8	694	14	BM984220	BM984220 UI-CF-EC1
22	21	2.8	698	9	AL563746	AL563746 AL563746
23	21	2.8	714	14	BM971307	BM971307 UI-CF-EC1
24	21	2.8	717	9	AA524013	AA524013 ng32304.s
25	21	2.8	723	12	RG740421	RG740421 602633809
26	21	2.8	750	9	AL539298	AL539298 AL539298
27	21	2.8	769	9	AL539297	AL539297 AL539297
28	21	2.8	769	9	AL539340	AL539340 AL539340
29	21	2.8	811	9	AL540037	AL540037 AL540037
30	21	2.8	1518	17	AG041349	AG041349 Pan-trogl
31	20	2.7	129	12	BF985977	BF985977 CM4-CM008
32	20	2.7	157	14	T26922	T26922 ESTD19684.C
33	20	2.7	188	10	AW867490	AW867490 MRO-SN003
34	20	2.7	189	14	W32760	W32760 zc06ell.r1
35	20	2.7	208	13	BG943498	BG943498 ax38e08.x
36	20	2.7	226	9	AI620074	AI620074 tv48b07.x
37	20	2.7	233	9	AI572419	AI572419 tp52e01.x
38	20	2.7	254	9	AA640492	AA640492 nt69a08.s
39	20	2.7	282	9	AA486925	AA486925 ab17b05.f
40	20	2.7	282	12	BG007728	BG007728 RC4-GN013
41	20	2.7	289	9	AA322055	AA322055 EST24631
42	20	2.7	306	17	AQ114103	AQ114103 CIT-HSP-2
43	20	2.7	318	12	BF799126	BF799126 CM2-CI003
44	20	2.7	322	14	T72674	T72674 yd19h01.s1
45	20	2.7	328	12	BF799093	BF799093 CM2-CI003

ALIGNMENTS

RESULT 1
AJ003597/c 338 bp mRNA linear EST 04-DEC-1997
LOCUS AJ003597 Selected chromosome 21 cDNA library Homo sapiens cDNA
DEFINITION clone MP19-12J9, mRNA sequence.
ACCESSION AJ003597
VERSION AJ003597.1 GI:2578270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 338)
AUTHORS Szulzewsky, I., Hunt, E., Nguyen, M., Korn, B., Roehrdanz, B., Lehrach, H. and Yaspo, M.L.H.
TITLE An integrated transcript map for the whole human chromosome 21
JOURNAL Unpublished (1997)
COMMENT Contact: Yaspo, M.-L.
Max Planck Institut fuer Molekulare Genetik
Innestrassse 73, D14195 Berlin-Dahlem, Germany.

FEATURES
source
1. 338

Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21g"
/clone="MP19-12J9"
/clone_lib="Selected chromosome 21 cDNA library"
/note="Swaroop et al. (1991) Nucleic Acids Res. 19: 1954."
BASE COUNT 76 a 92 c 82 g 87 t 1 others

QY	158	TCCTAGCATAAACTAGACACAT	190
Db	165	TCCTAGCATAAACTAGACACAT	187

RESULT 4
BG152895 linear EST 05-FEB-2001
LOCUS nad09h06.x1 NC1_CGAP_P128 Homo sapiens cDNA clone IMAGE:343187 3,
DEFINITION similar to contains L1.t2 L1 repetitive element ; , mRNA sequence.
ACCESSION BG152895

```

VERSION      BG152895.1  GI:12664925
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 432)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgabs-r@mail.nih.gov
             Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             DNA Sequencing by: Greg Lennon, Ph.D.
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             info@image.llnl.gov
             Seq primer: -40UP from Gibco
             High quality sequence stop: 393.
             Location/Qualifiers
               1..432
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:343187"
                 /clone_lib="NCI-CGAP_Pr28"
                 /sex="male"
                 /dev_stage="adult"
                 /lab_host="DH10B"
                 /note="Organ: prostate; Vector: pT7T3B-Pac (Pharmacia)
                 with a modified polylinker; Plasmid DNA from the
                 normalized library NCI-CGAP_Pr22 was prepared, and ss
                 circles were made in vitro. Following RAP purification,
                 this DNA was used as tracer in a subtractive hybridization
                 reaction. The driver was PCR-amplified cDNAs from a pool
                 of 5,000 clones made from the same library (clonoids
                 985608-986759, 1101192-1101959, and 1217928-1220615).
                 Subtraction by Bento Soares and M. Fatima Bonaudo."
BASE COUNT   179 a 90 c 78 g 85 t
ORIGIN
Query Match      3.1%; Score 23; DB 12; Length 432;
Best Local Similarity 100.0%; Pred.No. 0.76;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TCCTAGCATAAACTAGACACAT 190
      |||||
Db 384 TCCTAGCATAAACTAGACACAT 406

RESULT 5
BG965086 636 bp mRNA linear EST 12-JUN-2001
LOCUS 602829157F1 NCI-CGAP_Co24 Mus musculus cDNA clone IMAGE:4983830 5',
DEFINITION mRNA sequence.
ACCESSION BG965086
VERSION BG965086.1 GI:14352723
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 636)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgabs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.

FEATURES
source
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10989 row: a column: 15
High quality sequence start: 17
High quality sequence stop: 634.
Location/Qualifiers
  1..636
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:4983830"
    /clone_lib="NCI-CGAP_Co24"
    /lab_host="DH10B (TI phage-resistant)"
    /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
    Average insert size 1.6 kb. Constructed by Life
    Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT   155 a 172 c 155 g 154 t
ORIGIN
Query Match      3.1%; Score 23; DB 13; Length 636;
Best Local Similarity 100.0%; Pred.No. 0.84;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 CTGACCCCTGAAGACCATGATGGG 647
      |||||
Db 611 CTGACCCCTGAAGACCATGATGGG 633

RESULT 6
BG913391 641 bp mRNA linear EST 05-JUN-2001
LOCUS 602812247F1 NCI-CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944202
DEFINITION 5', mRNA sequence.
ACCESSION BG913391
VERSION BG913391.1 GI:14293867
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 641)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgabs-r@mail.nih.gov
             Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10888 row: n column: 11
             High quality sequence stop: 570.
             Location/Qualifiers
               1..641
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:4944202"
                 /clone_lib="NCI-CGAP_Brn67"
                 /tissue_type="anaplastic oligodendroglioma with lp/19q
                 loss"
                 /lab_host="DH10B (TI phage-resistant)"
                 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                 Average insert size 2.3 kb. Constructed by Life
                 Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT   173 a 171 c 167 g 130 t

```

ORIGIN

Query Match 3.1%; Score 23; DB 13; Length 641;
 Best Local Similarity 100.0%; Pred. No. 0.84;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TCCTAGCATAAACTAGACACAT 190 GSS 17-JUN-1999
 RPCI-11-462N4.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-462N4,
 DNA sequence.
 Db 143 TCCTAGCATAAACTAGACACAT 165

RESULT 7
 A0634393
 LOCUS
 DEFINITION
 A0634393 674 bp DNA linear GSS 17-JUN-1999
 RPCI-11-462N4.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-462N4,
 DNA sequence.
 VERSION
 A0634393.1 GI:5097028
 KEYWORDS
 GSS.
 SOURCE
 human.
 ORGANISM
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 674)

REFERENCE
 AUTHORS
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter

TITLE
 Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL
 COMMENT
 Map Building
 Unpublished (1997)

Other GSSs: RPCI-11-462N4.TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genet cs (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/humden/bac_end_search/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Source

Location/Qualifiers
 1..674
 /organism="Homo sapiens"
 /db_xref="GDB:7677339"
 /db_xref="taxon:9606"
 /clone="RPCI-11-462N4"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 194 a 172 c 166 g 141 t
 ORIGIN

Query Match 3.1%; Score 23; DB 17; Length 674;
 Best Local Similarity 100.0%; Pred. No. 0.85;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 TCCTAGCATAAACTAGACACAT 190
 Db 232 TCCTAGCATAAACTAGACACAT 254

RESULT 8
 BB284950
 LOCUS
 DEFINITION
 BB284950 286 bp mRNA linear EST 09-JUL-2000
 BB284950 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA
 clone B020002J03 3' similar to M77167 Mouse T-cell antigen receptor

alpha-chain (TCR-ATF2) mRNA, mRNA sequence.

BB284950
 BB284950.1 GI:8985399
 EST.
 house mouse.
 SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 286)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.

, Matsuyama, T., Miki, R., Mizuno, F., Nakamura, M., Oda, H., Okazaki, Y.

, Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Soabe, Y., Sugahara, Y.

, Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya

, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.

, Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino

, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki

, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki

, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

FEATURES

Source

Location/Qualifiers

1..286

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="B020002J03"

/clone_lib="RIKEN full-length enriched, 2 cells egg"

/tissue_type="egg"

/dev_stage="2 calls"

/lab_host="DH10B"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGATCCAGATCTTTTTTTTTTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGATCTCGATTAATAATATCCCCCCCCCCCCC 3']. cDNA

was cleaved with XhoI and BamHI. Vector: a modified

pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 86 a 63 c 44 g 93 t
ORIGIN

```

Query Match      2.8%; Score 21; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 AAGAAAGACGGGAAGCTTCCTT 369
      |||||||
DB 93 AAGAAAGACGGGAAGCTTCCTT 113

RESULT 9
AW901629
LOCUS      293 bp      mRNA      linear      EST 24-MAY-2000
DEFINITION RCO-NN1015-040400-021-e05 NN1015 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW901629
VERSION     AW901629.1 GI:8065834
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 293)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: -55-11-2704922
Fax: -55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t3-RCO-NN1015-040
400-021-e05&t3=2000-04-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 293.
Location/Qualifiers
1..293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1015"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      59 a 76 c 73 g 85 t
ORIGIN

Query Match      2.8%; Score 21; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 CTCCTCCACCTCAGCCTCCTA 172
      |||||||
DB 102 CTCCTCCACCTCAGCCTCCTA 122

RESULT 10
AW901629
LOCUS      293 bp      mRNA      linear      EST 24-MAY-2000
DEFINITION RCO-NN1015-040400-021-e05 NN1015 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW901629
VERSION     AW901629.1 GI:8065834
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 293)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: -55-11-2704922
Fax: -55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t3-RCO-NN1015-040
400-021-e05&t3=2000-04-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 293.
Location/Qualifiers
1..293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1015"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      59 a 76 c 73 g 85 t
ORIGIN

```

```

AI368142      325 bp      mRNA      linear      EST 13-FEB-1999
LOCUS      qq44907.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:1935420 3' similar to contains Alu repetitive element.; mRNA
sequence.
ACCESSION  AI368142
VERSION     AI368142
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 325)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1976 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 301.
Location/Qualifiers
1..325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      99 a 67 c 62 g 97 t
ORIGIN

Query Match      2.8%; Score 21; DB 9; Length 325;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 CTCCTCCACCTCAGCCTCCTA 172
      |||||||
DB 84 CTCCTCCACCTCAGCCTCCTA 104

RESULT 11
AQ368646
LOCUS      497 bp      DNA      linear      GSS 06-MAR-1999
DEFINITION HS_5038_A2_E10_77 RPC111 Human Male BAC Library Homo sapiens
genomic clone Plate=614 Col=20 Row=I, DNA sequence.
ACCESSION  AQ368646
VERSION     AQ368646
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 497)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

```


99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 614 Row: I Column: 20
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 497.
 Location/Qualifiers
 1. .497
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RPC111 Human Male BAC Library"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; RPC111 Human Male BAC Library"

BASE COUNT 107 a 118 c 90 g 177 t 5 others

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 497;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 TCTCCACCTCAGCCTCCTAG 173
 |||||||

Db 376 TCTCCACCTCAGCCTCCTAG 396
 |||||||

RESULT 12
 AW118817/c
 LOCUS
 DEFINITION x097a09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:2505528 3', mRNA sequence.

ACCESSION AW118817
 VERSION AW118817.1 GI:6087401
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 512)

REFERENCE
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 433.
 Location/Qualifiers
 1. .512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP-GCBI) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo.
 BASE COUNT 179 a 102 c 98 g 133 t

ORIGIN

Query Match 2.8%; Score 21; DB 10; Length 512;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 CTCCTCCACCTCAGCCTCCTA 172
 |||||||

Db 496 CTCCTCCACCTCAGCCTCCTA 476
 |||||||

RESULT 13
 BM967143
 LOCUS
 DEFINITION i31a05.v1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA clone IMAGE:6136209 5', mRNA sequence.

ACCESSION BM967143
 VERSION BM967143.1 GI:19560851
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 523)

REFERENCE
 AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blisstein, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagarisvilli, R., Williams, T.,
 Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other_ESTS: i31a05.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohpc.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brownjfas.harvard.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 447.
 Location/Qualifiers
 1. .523
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:6136209"
 /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
 /sex="Both"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
 Site_2: Sal 1; Starting library constructed using
 Superscript Plasmid library kit (Life Technologies). cDNA
 made by oligo-dr priming. Size-selected by column
 fractionation; average insert size 1.08 kb. Library was
 amplified once on solid support and plasmid DNA from
 library was prepared. The library DNA was normalized by
 method 44 from Bonaldo, Lennon, and Soares 1996 genome
 Research 6:791-806; 0.5 microgram single-stranded library
 plasmid DNA was mixed with 5 micrograms PCR product
 representing library inserts and hybridized to an EcoT of

20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 115 a 115 c 113 g 180 t
ORIGIN

Query Match 2.8%; Score 21; DB 14; Length 523;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 CTCCTCCACCTCAGCCTCCTA 172
|||||
Db 138 CTCCTCCACCTCAGCCTCCTA 158

RESULT 14

BM967160

LOCUS

BM967160 523 bp mRNA linear EST 29-APR-2002
l31bl1.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
cDNA clone IMAGE:6136269 5', mRNA sequence.

ACCESSION

BM967160

VERSION

BM967160.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 523)

Melton D., Brown J., Kenty G., Permutt A., Lee C., Kaestner K.,

Lemishka I., Secare M., Brestelli J., Gradwohl G., Clifton S.,

Hillier D., Marra M., Pape B., Wylie T., Martin J., Blisstein A.,

Schmitt A., Theising B., Ritter E., Ronko I., Bennett J., Cardenas

M., Gibbons M., McCann R., Cole R., Tsagarishvili R., Williams T.,

Jackson Y. and Bowers Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Juliana Brown

(brown@fas.harvard.edu)

Seq primer: -40RP from Gibco

High quality sequence stop: 446.

Location/Qualifiers

1..523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6136269"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site 1; Not 1;

Site 2: Sal 1; Starting library constructed using

SuperScript Plasmid Library kit (Life Technologies). cDNA

made by oligo-dT priming. Size-selected by column

fractionation; average insert size 1.08 kb. Library was

amplified once on solid support and plasmid DNA from

method #4 from Bonaldo, Lennon, and Soares 1996 Genome

Research 6:791-806; 0.5 microgram single-stranded library

Plasmid DNA was mixed with 5 micrograms PCR product

representing library inserts and hybridized to an Ecot of

20. Single-stranded (unhybridized) plasmids were isolated

by hydroxyapatite chromatography and used to make this

library."

Library."

Library."

Library."

Library."

Library."

BASE COUNT 115 a 115 c 112 g 181 t
ORIGIN

Query Match 2.8%; Score 21; DB 14; Length 523;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 CTCCTCCACCTCAGCCTCCTA 172
|||||
Db 138 CTCCTCCACCTCAGCCTCCTA 158

RESULT 15

AI281699

LOCUS

AI281699

DEFINITION

AI281699

ACCESSION

AI281699

VERSION

AI281699.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 537)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip

Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 707 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 325.

Location/Qualifiers

1..537

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1960365"

/clone_lib="NCI-CGAP_Eso2"

/tissue_type="squamous cell carcinoma"

/lab_host="DH10B"

/note="Organ: esophagus; Vector: pCMV-SPORT6; Site 1: SalI

; Site 2: NotI; Cloned unidirectionally. Primer: Oligo

dT. Average insert size 1.1 kb. Life technologies catalog

#: 11502-010"

BASE COUNT 128 a 132 c 100 g 175 t

ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 537;

Best Local Similarity 100.0%; Pred. No. 9.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 TCTCCACCTCAGCCTCCTAG 173

|||||

Db 135 TCTCCACCTCAGCCTCCTAG 155

Search completed: April 12, 2003, 08:59:20
Job time : 831.43 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 03:59:25 ; Search time 20.8447 Seconds
(without alignments)
11049.076 Million cell updates/sec

Title: US-09-914-152-3_COPY_8000_8750

Perfect score: 751
Sequence: 1 ggaccatgcccagaatctc.....gactgaactgtcttctgaaga 751

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 1533381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	19	2.5	449	2	US-08-967-101-75
2	19	2.5	449	2	US-08-592-541-75
3	19	2.5	449	3	US-09-124-698-75
4	19	2.5	449	4	US-09-127-480-75
5	19	2.5	449	4	US-08-496-841C-75
6	19	2.5	449	4	US-09-124-523-75
7	19	2.5	500	2	US-08-967-101-109
8	19	2.5	500	2	US-08-592-541-109
9	19	2.5	500	3	US-09-124-698-109
10	19	2.5	500	4	US-09-127-480-109
11	19	2.5	500	4	US-08-496-841C-109
12	19	2.5	500	4	US-09-124-523-109
13	19	2.5	599	4	US-09-222-575-173
14	19	2.5	633	4	US-08-325-111-558
15	19	2.5	945	2	US-08-967-101-6
16	19	2.5	945	2	US-08-967-101-160
17	19	2.5	945	2	US-08-592-541-6
18	19	2.5	945	2	US-08-592-541-160
19	19	2.5	945	3	US-08-888-077A-10
20	19	2.5	945	3	US-09-124-698-6
21	19	2.5	945	3	US-09-124-698-160
22	19	2.5	945	4	US-09-127-480-6
23	19	2.5	945	4	US-09-127-480-160
24	19	2.5	945	4	US-08-496-841C-6
25	19	2.5	945	4	US-08-496-841C-160
26	19	2.5	945	4	US-09-124-523-6
27	19	2.5	945	4	US-09-124-523-160

c 28 19 2.5 1001 4 US-09-641-638-526 Sequence 526, App
c 29 19 2.5 4866 1 US-08-110-158-5 Sequence 5, Appl
c 30 19 2.5 18073 4 US-09-078-294-12 Sequence 12, Appl
c 31 19 2.5 26664 4 US-09-564-805-28 Sequence 28, Appl
c 32 19 2.5 28720 4 US-09-341-587-7 Sequence 7, Appl
c 33 19 2.5 35060 3 US-08-814-095-7 Sequence 7, Appl
c 34 19 2.5 35100 1 US-08-306-691B-19 Sequence 19, Appl
c 35 19 2.5 35100 5 PCT-US93-06251-19 Sequence 19, Appl
c 36 19 2.5 56516 2 US-08-996-306-1 Sequence 1, Appl
c 37 19 2.5 56516 4 US-09-338-907-1 Sequence 1, Appl
c 38 19 2.5 56516 4 US-09-218-207-1 Sequence 1, Appl
c 39 19 2.5 56520 4 US-09-338-907-179 Sequence 179, App
c 40 19 2.5 56520 4 US-09-218-207-179 Sequence 179, App
c 41 19 2.5 72604 4 US-09-268-992-7 Sequence 7, Appl
c 42 19 2.5 72604 4 US-09-657-474-7 Sequence 7, Appl
c 43 19 2.5 81001 4 US-09-750-580-1 Sequence 1, Appl
c 44 19 2.5 99500 4 US-09-798-096-10 Sequence 10, Appl
c 45 19 2.5 162450 4 US-09-345-882-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-967-101-75
; Sequence 75, Application US/08967101
; Patent No. 5840540
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURNITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,101
; FILING DATE: 10-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-967-101-75

Query Match 2.5%; Score 19; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCCTCA 172
|||||

```
Db 33 CTCACCTCAGCCTCCTA 51
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-124-698-75
;
Query Match 2.5% Score 19; DB 3; Length 449;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 154 CTCACCTCAGCCTCCTA 172
|||||
Db 33 CTCACCTCAGCCTCCTA 51
;
;
RESULT 4
US-09-127-480-75
; Sequence 75, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: DNA (genomic)
; US-09-127-480-75
;
Query Match 2.5% Score 19; DB 2; Length 449;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 154 CTCACCTCAGCCTCCTA 172
|||||
Db 33 CTCACCTCAGCCTCCTA 51
;
;
RESULT 3
US-09-124-698-75
; Sequence 75, Application US/09124698~
; Patent No. 6117978
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-75

Query Match 2.5%, Score 19; DB 4; Length 449;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCTCCTTA 172
DB 33 CTCGCCACCTCAGCCTCCTTA 51

RESULT 5
US-08-496-841C-75
; Sequence 75, Application US/08456841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; ROMMENS, JOHANNA M
; FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-496-841C-75
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/496.841C
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-124-523-75

Query Match 2.5%, Score 19; DB 4; Length 449;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCTCCTTA 172
DB 33 CTCGCCACCTCAGCCTCCTTA 51

RESULT 6
US-09-124-523-75
; Sequence 75, Application US/09124523
; Patent No. 6395960
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; ROMMENS, JOHANNA M
; FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967.101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:

LENGTH: 500 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-967-101-109

Query Match 2.5%; Score 19; DB 2; Length 500;

Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172

|||||

Db 33 CTCGCCCTCAGCCTCCTA 51

RESULT 8

US-08-592-541-109

Sequence 109, Application US/08592541

Patent No. 5986054

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:

LENGTH: 500 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-592-541-109

Query Match 2.5%; Score 19; DB 2; Length 500;

Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172

|||||

Db 33 CTCGCCCTCAGCCTCCTA 51

RESULT 9

US-09-124-698-109

Sequence 109, Application US/09124698

Patent No. 6117978

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/124,698

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:

LENGTH: 500 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-124-698-109

Query Match

Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172

|||||

Db 33 CTCGCCCTCAGCCTCCTA 51

RESULT 10

US-09-127-480-109

Sequence 109, Application US/09127480

Patent No. 6194153

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TITLE OF INVENTION: TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127.480
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592.541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: linear
TOPOLOGY: linear
SEQUENCE TYPE: DNA (genomic)
US-09-127-480-109

Query Match 2.5%; Score 19; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCTCCTA 172
Db 33 CTCGCCACCTCAGCCTCCTA 51

RESULT 11
US-08-496-841C-109
Sequence 109, Application US/08496841C
Patent No. 6210919
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496.841C
FILING DATE: 28-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul F. Fehlnner, Ph.D.
REGISTRATION NUMBER: 35,135
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
```

```
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-08-496-841C-109

Query Match 2.5%; Score 19; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCTCCTA 172
Db 33 CTCGCCACCTCAGCCTCCTA 51

RESULT 12
US-09-124-523-109
Sequence 109, Application US/09124523
Patent No. 6395960
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124.523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592.541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-124-523-109

Query Match 2.5%; Score 19; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCACCTCAGCCTCCTA 172
Db 33 CTCGCCACCTCAGCCTCCTA 51

RESULT 13
US-09-222-575-173
Sequence 173, Application US/09222575
Patent No. 6387697
GENERAL INFORMATION:
APPLICANT: YUQIU, JIANG
```

APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
TITLE OF INVENTION: and Methods for Their Use
FILE REFERENCE: 210121.470
CURRENT APPLICATION NUMBER: US/09/222.575
CURRENT FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 174
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 173
LENGTH: 599
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: modified_base
LOCATION: (5)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (31)
OTHER INFORMATION: Where n is a, c, g or t
US-09-222-575-173

Query Match 2.5%; Score 19; DB 4; Length 599;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172
DB 542 CTCGCCCTCAGCCTCCTA 560

RESULT 14

US-09-328-111-358/C
Sequence 358, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Cartoll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1998-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 358
LENGTH: 633
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(633)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-358

Query Match 2.5%; Score 19; DB 4; Length 633;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172

DB 325 CTCGCCCTCAGCCTCCTA 307

RESULT 15

US-08-967-101-6/C
Sequence 6, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMKENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 945 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-967-101-6

Query Match 2.5%; Score 19; DB 2; Length 945;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 CTCGCCCTCAGCCTCCTA 172
DB 913 CTCGCCCTCAGCCTCCTA 895

Search completed: April 12, 2003, 09:03:16
Job time : 25.8447 secs

Db 121 GACGGAACTTCCTTAAGCTCCAGATACAGACTGCAGCAGACACCTCCCTCTCTGTC 180
 QY 182 CTGCTGGTGCCTCATCCCAACAGTTGGCTGAGCGCATGCCATCGGACAGCTGG 241
 Db 181 CTGCTGGTGCCTCATCCCAACAGTTGGCTGAGCGCATGCCATCGGACAGCTGG 240
 QY 242 GGAAGAAGAGAGTGGTGAAGGAAAGAGCTGAGACATTTCTCTCTCTGGGACACC 301
 Db 241 GGAAGAAGAGAGTGGTGAAGGAAAGAGCTGAGACATTTCTCTCTCTGGGACACC 300
 QY 302 AGCAGTGCAGCGGAAACGAAGAGTGGACAGGAGGACGAGCAGCGACGAGATATC 361
 Db 301 AGCAGTGCAGCGGAAACGAAGAGTGGACAGGAGGACGAGCAGCGAGATATC 360
 QY 362 CAGAAGAGATTCCTAGACGCTATTACATCTGACCTGAAAGACATGATGGCATAGAA 421
 Db 361 CAGAAGAGATTCCTAGACGCTATTACATCTGACCTGAAAGACATGATGGCATAGAA 420
 QY 422 TGGGTCCATCGCTTTTGTCTCAGCGGCGCTTTGTGATGAAACAGACTCAGACATGTC 481
 Db 421 TGGGTCCATCGCTTTTGTCTCAGCGGCGCTTTGTGATGAAACAGACTCAGACATGTC 480
 QY 482 ATCAATCTGACTATCTGACTGACTGCTCTGAGAAACAGACACAGCTTTTTC 541
 Db 481 ATCAATCTGACTATCTGACTGACTGCTCTGAGAAACAGACACAGCTTTTTC 540
 QY 542 ACTGCTCTTTGAAACTCAATGAGTTTCCCATCAGGAGGACATTCAGCAAGTGTTC 601
 Db 541 ACTGCTCTTTGAAACTCAATGAGTTTCCCATCAGGAGGACATTCAGCAAGTGTTC 600
 QY 602 AGTAATCTGAATATCCGTGGGAGGAGTACCCACCTTTCTGCTCCGCGACCGGCTACGTG 661
 Db 601 AGTAATCTGAATATCCGTGGGAGGAGTACCCACCTTTCTGCTCCGCGACCGGCTACGTG 660
 QY 662 TTTTCTGGGAGCGGCGAGTCAGTGTACAAATGTCTCCAGAGCGTCCCATACATTA 721
 Db 661 TTTTCTGGGAGCGGCGAGTCAGTGTACAAATGTCTCCAGAGCGTCCCATACATTA 720
 QY 722 CTGGAAGAGCTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 781
 Db 721 CTGGAAGAGCTTTGTGGGCTCTGCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 780
 QY 782 CACTCCCGAGCGCTTTTTCAGGGGCTTAGCTTCTCCGATGCTCTCAGGAGG 841
 Db 781 CACTCCCGAGCGCTTTTTCAGGGGCTTAGCTTCTCCGATGCTCTCAGGAGG 840
 QY 842 ATCTGGGCTGCTCCTCATCAAGCTCGGACTCTCTTGAGTACTTGGAGGCTCTAGAG 901
 Db 841 ATCTGGGCTGCTCCTCATCAAGCTCGGACTCTCTTGAGTACTTGGAGGCTCTAGAG 900
 QY 902 AATTCGGGGGGAAGATTGTCGCGCTGC 931
 Db 901 AATTCGGGGGGAAGATTGTCGCGCTGC 930

RESULT 4
 E38420
 LOCUS E38420 10562 bp DNA linear PAT 31-JAN-2002
 DEFINITION Novel polypeptide.
 ACCESSION E38420
 VERSION E38420.1 GI:18626994
 KEYWORDS JP 2000245464-A/2.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
 TITLE Novel polypeptide.
 JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;
 KYOWA HAKKO KOGYO CO LTD
 OS Homo sapiens (human)
 PN JP 2000245464-A/2

REFERENCE
 AUTHORS Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
 TITLE Novel polypeptide.
 JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;
 KYOWA HAKKO KOGYO CO LTD
 OS Homo sapiens (human)
 PN JP 2000245464-A/2

REFERENCE
 AUTHORS Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
 TITLE Novel polypeptide.
 JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;
 KYOWA HAKKO KOGYO CO LTD
 OS Homo sapiens (human)
 PN JP 2000245464-A/2

REFERENCE
 AUTHORS Narimatsu, H., Isshiki, S., Togayauchi, A. and Sasaki, K.
 TITLE Novel polypeptide.
 JOURNAL Patent: JP 2000245464-A 2 12-SEP-2000;
 KYOWA HAKKO KOGYO CO LTD
 OS Homo sapiens (human)
 PN JP 2000245464-A/2

PD 12-SEP-2000
 PF 25-FEB-1999 JP 1999047571
 PR HISASHI NARIMATSU, SOICHIRO ISSHIKI, AKIRA TOGAYAUCHI, PI
 KATSUOSHI SASAKI
 PC C12N15/09, A01K67/027, C12N1/21, C12N5/10, C12N9/10, C12P19/00, PC
 C12P21/02, C12Q1/68, G01N33/53, C12N1/21, C12R1/185, C12N5/10,
 C12P21/02, C12R1/185, C12P21/02, C12R1/91, C12N15/00, C12N5/00,
 C12P21/02, C12R1/91
 CC Key Location/Qualifiers
 FT Promoter (1), (5000)
 FT Exon (5001) .. (5140)
 FT Exon (5001) .. (5273)
 FT Exon (5459) .. (5567)
 FT Exon (7427) .. (7586)
 FT Exon (8234) .. (10562)
 FEATURES
 source
 1. 10562
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 2610 a 2415 c 2574 g 2963 t
 ORIGIN
 Query Match 99.9%; Score 930; DB 6; Length 10562;
 Best Local Similarity 100.0%; Pred. No. 3; le-242;
 Matches 930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ATGGCTTTCCGGAAGATGAGATTGATGATATATTTGCTTCTGCTTCTGGGGCTCTTTGT 61
 Db 8234 ATGGCTTTCCGGAAGATGAGATTGATGATATTTGCTTCTGCTTCTGGGGCTCTTTGT 8293
 QY 62 TTGTATTTAGCATGTACAGTCTAAATCCCTTCAAGACAGACGCTTTGTACAGAAA 121
 Db 8294 TTGTATTTAGCATGTACAGTCTAAATCCCTTCAAGACAGACGCTTTGTACAGAAA 8353
 QY 122 GACGGAACTCTCTTAAGCTCCAGATACAGACTGCGAGGAGACACCTCCCTTCCTCGTC 181
 Db 8354 GACGGAACTCTCTTAAGCTCCAGATACAGACTGCGAGGAGACACCTCCCTTCCTCGTC 8413
 QY 182 CTGCTGGTGCCTCTATCCCAACAGTTGGCTGAGCGCATGCCATCGGACAGAGTGG 241
 Db 8414 CTGCTGGTGCCTCTATCCCAACAGTTGGCTGAGCGCATGCCATCGGACAGAGTGG 8473
 QY 242 GGAAGAGAGGATGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGACACC 301
 Db 8474 GGAAGAGAGGATGGTGAAGGAAAGCAGCTGAAGACATTTCTCTCTGGGACACC 8533
 QY 302 AGCAGTGCAGGAAACGAAAGAGTGGACAGGAGAGCGAGACGAGCGGACATTATC 361
 Db 8534 AGCAGTGCAGGAAACGAAAGAGTGGACAGGAGAGCGAGACGAGCGGACATTATC 8593
 QY 362 CAGAAGGATTTCTAGAGCTCTATTACAATGTGACCTGAAGACCATGATGGGCATAGAA 421
 Db 8594 CAGAAGGATTTCTAGAGCTCTATTACAATGTGACCTGAAGACCATGATGGGCATAGAA 8653
 QY 422 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTCTGATGAAGACAGACTCAGACATGTC 481
 Db 8654 TGGGTCCATCGCTTTTGTCTCAGCGCGCTTCTGATGAAGACAGACTCAGACATGTC 8713
 QY 482 ATCAATGTTGACTATCTGACTGAACTGCTCTGAAGAAAAACAGAACACAGGTTTTC 541
 Db 8714 ATCAATGTTGACTATCTGACTGAACTGCTCTGAAGAAAAACAGAACACAGGTTTTC 8773
 QY 542 ACTGGCTCTTCAAACTCAATGAGTTTCCCATCAGGAGCGCATTCAGCAAGTGTGTTC 601
 Db 8774 ACTGGCTCTTCAAACTCAATGAGTTTCCCATCAGGAGCGCATTCAGCAAGTGTGTTC 8833
 QY 602 AGTAATCTGAATATCCGTGGGACAGGTACCCACCTTCTGCTCCGCGACCGGTACGTG 661
 Db 8834 AGTAATCTGAATATCCGTGGGACAGGTACCCACCTTCTGCTCCGCGACCGGTACGTG 8893

841 CATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTTTGGACTACTGGCAGGCTCTAGA 900
QY
1241 GATCGTGGCGCTGCCACTTCATCAAGCCTCGGACTCTTTGGACTACTGGCAGGCTCTAGA 1300
DB

Qy 901 GAATTCGGGGGAAGATTGTCCGGCTGC 931
 |||||
 Db 1301 GAATTCGGGGGAAGATTGTCCGGCTGC 1331

RESULT 3
AF145784

DEFINITION	Homo sapiens beta1,3 galactosyltransferase-V (B3GALT5) gene, complete cds.
ACCESSION	AF145784
VERSION	AF145784.1 GI:6409192

KEYWORDS	Homo sapiens.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 933)
AUTHORS	Zhou,D., Berger,E.G. and Hennet,T.
TITLE	Molecular cloning of a human UDP-galactose:GlcNAc6beta1,3galNAc beta1,3 galactosyltransferase gene encoding an O-linked cores-elongation enzyme
JOURNAL	Eur. J. Biochem. 263 (2), 571-576 (1999)
MEDLINE	99337698
PUBMED	10406968
REFERENCE	2 (bases 1 to 933)
AUTHORS	Zhou,D. and Hennet,T.
TITLE	Direct Submission
JOURNAL	Submitted (26-APR-1999) Physiology, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland
FEATURES	Location/Qualifiers
	1..933

```

gene      /map=21942
          /gene="B3GALT5"
          /gene="B3GALT5"
          /note="glycosyltransferase"
          /codon_start=1
          /product="beta1,3 galactosyltransferase-V"
          /protein_id="RAF07880.1"
          /db_xref="GI:6408193"
          /translation="MAFPRMLMYICLLVYGALCLIFSMYSLNPKFESFYVKKDGNF
          LKLPDQCRQPPPELVLLVTSYSHKQLAERMAIQWTGKRMVYKQQLAIFFDLGTSS
          AAETKEVQESQRGDIQKDFLDVYNLLKTMGIEWHFVPCQAAFAVNRKTDSDMF
          INVYDITELLAKNRTIRFTGLKLNFEPIRQPFKWFVSKEYPMWRYPYFCSGTG
          YVFSGDVAQVNYNVKSPVYIKLEDVFGVLGRRLNIRLELHQSOTFFPGGLRFSVC
          LFRRIVACHFIKPTLLDYQALNSRSGDCPV"
BASE COUNT      227 a 232 c 241 g 233 t
ORIGIN
Query Match      99.9%; Score 930; DB 9; Length 933;
Best Local Similarity 100.0%; Pred. No. 2,9e-242;
Matches 930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2  ATGCGCTTCCCGAAGATGAGATGATATTTGCCCTTCGTGGTCTCGGGGGCTCTTTGT 61
Ddb      1  ATGCGCTTCCCGAAGATGAGATGATATTTGCCCTTCGTGGTCTCGGGGGCTCTTTGT 60
QY      62  TTGTATTTAGCATGTACAGTCTAAATCCTTCAAGAAGACAGTCCCTTTGTTTACAAGAA 121
Ddb      61  TTGTATTTAGCATGTACAGTCTAAATCCTTCAAGAAGACAGTCCCTTTGTTTACAAGAA 120
QY      122 GACGGGAACCTTCCTTAAGCTCCACAGATACAGACTCAGGAGACAGACCTCCCTTCCTCGTC 181
Ddb      122 GACGGGAACCTTCCTTAAGCTCCACAGATACAGACTCAGGAGACAGACCTCCCTTCCTCGTC 181

```


Qy 841 GATCGTGGCGCTGGCACTTCATCAAGCTCGGACTCTCTTGACTACTGGCAGGCTCTAGA 900

D8 1241 GATCGTGGCGCTGGCACTTCATCAAGCTCGGACTCTCTTGACTACTGGCAGGCTCTAGA 1300

QY 901 GAATTCGGGGGGAAGATTGTCCGCGCTTC 931

Db 1301 GAATCCCGGGGAGATTGTCCGCTGTC 1331

RESULT 3

RESULT 3
 AF145784
 LOCUS
 033 bp
 DB145784
 140000
 DB1 14-NOV-2000

LOCUS	AF145784	DNA	linear	933 bp	PR1 14-NOV-1999
DEFINITION	Homo sapiens beta1,3 galactosyltransferase-V (B3GALT5) gene,				

ACCESSION	complete cds.
AF145784	

VERSION AF145784.1 GI:6409192
KEYWORDS

SOURCE	ORGANISM
Homo sapiens.	Homo sapiens
Homo sapiens	Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
1 (bases 1 to 933)
Zhou, D., Berger, E.G. and Hennet, T.
AUTHORS

TITLE Molecular cloning of a human UDP-galactose:GlcNAc6P,3GalNAc6P, 3 galactosyltransferase gene encoding an O-linked

JOURNAL OF
BACTERIOLOGY
163(2):571-576 (1999)

EDITED BY
 J. J. BLOCH, 205 (2), 311-376 (1999)
 MEDLINE
 99337698
 10406058
 DISSEMINATED

PURIFIED
10406968
REFERENCE
2 (bases 1 to 933)
SEQUENCE INFORMATION

AUTHORS	TITLE	Direct Submission	Indirect Submission
Zhou, B. and Hennes, T.	Direct Submission	1000	1000

JOURNAL Submitted (26-APR-1999) Physiology, University of Zurich,
Winterthurerstrasse 190, Zurich 8057, Switzerland

FEATURES	Location/Qualifiers
source	1. .933

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

[/chromosome="21"](#)
[/map="21q22"](#)

```

gene
  /map_pos 1..933
  /gene="B3GALT5"

```

```

CDS
    /gene="B000000000"
    1..933
    /name="B303175"

```

```
/gene="BGGALD3"  
/note="glycosyltransferase"
```

```

/codon_start=1
/product="beta1,3 galactosyltransferase-V"
/feature_id="beta1,3 galactosyltransferase-V"

```

```
/protein_id="AAF07880.1"
/db_xref="GI:6409193"
```

```
/translation="MAFPKRWLMYICLLVLGALCLYFSNYSLNPPKQSFVYKKDGN  
LKLPTDCRQPPFLVLLVTSKKQLAERMAIRQTWGRNRMVKGOLKTLFFLLGTSS
```

AAETKEVDQESQRHGDIIQKDFLDVYYNLTLLKTMGIEVWHRFCQAFAFMKTDSDME
INVLYTELLLNKRNTRTEFTTGLKLNFFIRQPFSEKWEVSKSEYPWDRYPPFCSGTG

YVFGDVASQYNVSKSPYIKLEDVFGCLERLNI RLEELHSQPTFFPGGLRFSVC
LFRRIVACHFTKPRTLTDYQALNSRGEDCPV¹⁸

BASE COUNT	227 a	232 c	241 g	233 t
ORIGIN				

Query Match	99.9%	Score 930	DB 9	Length 933
ORIGIN				

Query Match	99.9%	Score 930	DB 3	Length 930
Best Local Similarity	100.0%	Pred. No. 2.9e-242		
Matches 930	Conservative	0	Mismatches 0	Indels 0
Gaps 0				

```

Matches 930; Conservative 0; Mismatches 0; Indels 0; Gaps

```

2 ATGGCTTCCCGAGATGAGATTGATATATTCCTTCCTGGGTCTCTGGGGCTCTTGT 61

Db 1 ATGGCTTCCCGAAGATGAGATTGATATTTGGCTTCTGGTCTCGGGGCTCTTTGT 60

Qy 62 TTTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGACACAGCTCTTTTGTTTACAGAAA 121

Db 61 TTGTATTTAGCATGTACAGTCTAAATCCTTCAAGAACAGTCCTTGTGTTACAGAAA 120

Qy 122 GACGGGAACCTCCTTAAGCTCCAGATACAGACTGGAGGCAGACACCTCCCTTCCTGTC 181

.....

1

100

RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [5]
 RA Sequence of 1-298 FROM N.A.
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP REVIEW.
 RX MEDLINE=20047730; PubMed=10580128;
 RA Amado M., Almeida R., Schwientek T., Clausen H.;
 RT "Identification and characterization of large galactosyltransferase
 RT gene families: galactosyltransferases for all functions.";
 RL Biochim. Biophys. Acta 1473:35-53(1999).
 CC [1]- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
 CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
 CC GLCNAC(BETAL.3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
 CC EFFICIENT ACCEPTOR.
 CC [1]- PATHWAY: GLYCOSYLATION.
 CC [1]- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC [1]- TISSUE SPECIFICITY: EXPRESSED IN STOMACH, JEJUNUM, COLON,
 CC PANCREAS, SMALL INTESTINE, TESTIS AND GASTROINTESTINAL AND
 CC PANCREATIC CANCER CELL LINES. HARDLY DETECTED IN LUNG, LIVER,
 CC ADRENAL GLAND AND PERIPHERAL BLOOD LEUKOCYTES.
 CC [1]- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AB020337; BAA77654.1; -
 CC EMBL: AF145784; AAF07880.1; -
 CC EMBL: AJ006078; CAB91547.1; -
 CC EMBL: AL163280; CAB90446.1; -
 CC EMBL: AB041412; BAA94497.1; -
 CC EMBL: AB041413; BAA94498.1; -
 CC Genbank: HGNC:920; B3GALT5.
 CC MIM: 604066; -
 CC InterPro: IPR002659; Galactosyl_T.
 CC Pfam: PF01762; Galactosyl_T.1.
 CC Transferase: Glycosyltransferase; Glycoprotein; Transmembrane;
 CC Signal-anchor; Golgi stack; Multigene family.
 CC DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 29 310 LUMENAL, CATALYTIC (POTENTIAL).
 CC CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CONFLICT 23 23 F -> Y (IN REF. 3).
 CC CONFLICT 26 26 Y -> N (IN REF. 3).
 CC CONFLICT 42 42 G -> V (IN REF. 5; BAA94497).
 CC CONFLICT 85 85 M -> T (IN REF. 3 AND 5; BAA94497).
 CC CONFLICT 114 114 K -> R (IN REF. 5; BAA94497).
 CC CONFLICT 234 234 K -> E (IN REF. 5; BAA94497).
 CC CONFLICT 277 277 L -> R (IN REF. 5; BAA94497).
 CC CONFLICT 293 297 LDTWO -> WTGR (IN REF. 5; BAA94498).
 CC SEQUENCE 310 AA; 36189 MW; 4DD7A19E3E648AA9 CRC64;
 Query Match 100.0%; Score 1657; DB 1; Length 310;
 Best Local Similarity 100.0%; Pred. No. 6.6e-146;
 Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAFPRMLWYICLVZGALCLYSMTSLNPFKEQSVYKDGNEFLKLPDDDCROTPEFLV 60
 DB 1 MAFPRMLWYICLVZGALCLYSMTSLNPFKEQSVYKDGNEFLKLPDDDCROTPEFLV 60
 QY 61 LVTSSHKQLAERMAIRQTKGRMVKGKOLKTFLLGTSSAAETKEVDQESQRGDII 120
 DB 1 LVTSSHKQLAERMAIRQTKGRMVKGKOLKTFLLGTSSAAETKEVDQESQRGDII 120

DB 61 LVTSSHKQLAERMAIRQTKGRMVKGKOLKTFLLGTSSAAETKEVDQESQRGDII 120
 QY 121 QKDFLDVYVNLTKTMGCIEMVHRFCPOAFAVMKTSDFINVDYLTLLAKNRTTRFF 180
 DB 121 QKDFLDVYVNLTKTMGCIEMVHRFCPOAFAVMKTSDFINVDYLTLLAKNRTTRFF 180
 QY 181 TGFLKLNFFPIRQPEFKWFSKSEYPMWDRYPFCSCGTGYVFGSDVASQVYVSKSPYIK 240
 DB 181 TGFLKLNFFPIRQPEFKWFSKSEYPMWDRYPFCSCGTGYVFGSDVASQVYVSKSPYIK 240
 QY 241 LEDVYVGLCLERLNRLLELSQTFPPGGLRFSVCLFRIVACHFIKRTLLDYWALE 300
 DB 241 LEDVYVGLCLERLNRLLELSQTFPPGGLRFSVCLFRIVACHFIKRTLLDYWALE 300
 QY 301 NSRGEDCPVP 310
 DB 301 NSRGEDCPVP 310
 RESULT 2
 B3G5_GORGO STANDARD; PRT; 300 AA.
 ID B3G5_GORGO
 AC Q9N293;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
 DE (betaGal-T5) (betaGal-T5) (UDP-galactose:beta-N-acetylgalactosamine beta-
 DE 1,3-galactosyltransferase 5) (UDP-gal:beta-GlcNAc beta-1,3-
 DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
 GN B3GALT5.
 OS Gorilla gorilla gorilla (Lowland gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
 OX NCBI_TaxID=9595;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Liu Y., Saitou N.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC [1]- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
 CC GLCNAC(BETAL.3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
 CC EFFICIENT ACCEPTOR (BY SIMILARITY).
 CC [1]- PATHWAY: GLYCOSYLATION.
 CC [1]- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC [1]- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AB041416; BAA94501.1; -
 CC InterPro: IPR002659; Galactosyl_T.
 CC Pfam: PF01762; Galactosyl_T.1.
 CC Transferase: Glycosyltransferase; Glycoprotein; Transmembrane;
 CC Signal-anchor; Golgi stack; Multigene family.
 CC DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 29 >300 LUMENAL, CATALYTIC (POTENTIAL).
 CC CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC NON_TER 300 300
 CC SEQUENCE 300 AA; 35104 MW; 1AA11692ED5F06FA CRC64;
 Query Match 96.1%; Score 1592; DB 1; Length 300;
 Best Local Similarity 99.7%; Pred. No. 6.5e-140;
 Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:22 ; Search time 14 seconds
(without alignments)
918.404 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 NAFPKRLMVICLLVGLAC.....TLLDYWQALENSRGEDCPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1657	100.0	310	1 B3G5_HUMAN	Q9Y2C3 h beta-1,3-
2	1592	96.1	300	1 B3G5_GORGO	Q9A293 g beta-1,3-
3	1570	94.7	301	1 B3G5_PANPA	Q9A294 p beta-1,3-
4	1552	93.7	297	1 B3G5_PANTR	Q9A295 p beta-1,3-
5	1206	72.8	308	1 B3G5_MOUSE	Q91167 m beta-1,3-
6	416	25.1	372	1 B3G8_HUMAN	Q9Y2A9 h probable
7	360	21.7	398	1 B3G7_HUMAN	Q9Y97 h beta-1,3-
8	105.5	6.4	1049	1 VP39_YEAST	Q07468 saccharomyc
9	93.5	5.6	336	1 MTHB_HAPPH	P50193 haemophilus
10	90.5	5.5	1314	1 S522_YEAST	P23390 saccharomyc
11	89.5	5.4	786	1 TLRL_HUMAN	Q15399 homo sapien
12	88.5	5.3	579	1 YHVO_YEAST	P38848 saccharomyc
13	86.5	5.2	516	1 YVNF_AZOCH	P24423 azotobacter
14	86.5	5.2	711	1 ETF2_MYXVL	Q9Q8K4 myxoma viru
15	85.5	5.2	632	1 Y242_MYCPN	P75440 mycoplasma
16	85	5.1	1073	1 HSER_PIG	P55204 sus scrofa
17	84.5	5.1	454	1 Y662_HAEIN	P45217 haemophilus
18	84	5.1	306	1 YFV9_SCHPO	Q9P7Q6 schizosacch
19	83	5.0	275	1 TSPH_THEMA	Q9X1F7 thermotoga
20	82.5	5.0	525	1 DIMH_CABEL	Q17397 caenorhabdi
21	82.5	5.0	713	1 PMIP_HUMAN	Q99797 homo sapien
22	82	4.9	576	1 HXT4_YEAST	P32467 saccharomyc
23	81.5	4.9	359	1 UNG_YEAST	P12887 saccharomyc
24	81.5	4.9	1042	1 SVI_BORBU	Q51773 borrelia bu
25	81	4.9	290	1 LEP_HELFP	Q9A2L5 helicobacte
26	81	4.9	397	1 O22A_DROME	P81909 drosophila
27	81	4.9	399	1 TRPB_BUCAP	P42391 buchera ap
28	81	4.9	1157	1 N133_YEAST	P31611 saccharomyc
29	80.5	4.9	290	1 LEP_HELFP	Q25300 helicobacte
30	80.5	4.9	711	1 EFT2_SFVKA	Q988Y2 snipe fibro
31	80.5	4.9	3386	1 POLG_DEN4	P03866 d genome po
32	79.5	4.8	612	1 BIR3_MOUSE	Q62210 mus musculu
33	79.5	4.8	691	1 Y104_YEAST	P40460 saccharomyc

34 78.5 4.7 757 1 METE_PASRU
35 78.5 4.7 1002 1 YA37_ANASP
36 78.5 4.7 4377 1 ANK3_HUMAN
37 78 4.7 469 1 LEU2_BUCAP
38 78 4.7 519 1 YG31_METJA
39 78 4.7 618 1 BIR3_HUMAN
40 78 4.7 741 1 GNT5_HUMAN
41 78 4.7 908 1 DPOL_BORBU
42 77.5 4.7 271 1 Y398_METJA
43 77.5 4.7 466 1 ARLY_CLOPE
44 77.5 4.7 710 1 PMIP_PAT
45 77 4.6 246 1 Y402_BUCAI

ALIGNMENTS

RESULT 1
B3G5_HUMAN STANDARD; PRT: 310 AA.
AC Q9Y2C3; Q9Y2A9; Q9Y2A9; Q9Y2A9; Q9Y2A9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (beta3gal-T5) (beta3gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 5) (Beta-3-Gx-T5)
GN B3GALT5
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Colorectal adenocarcinoma;
RX MEDLINE-99210289; PubMed-1021226;
RA Isshiki S., Toyachi A., Kudo T., Nishihara S., Watanabe M., Kubota T., Katsujima M., Shiraishi N., Sasaki K., Andoh T., Narimatsu H.;
RA "Cloning, expression, and characterization of a novel UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase (beta3gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived therefrom.";
RL J. Biol. Chem. 274:12499-12507(1999). April 99
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99337698; PubMed-10406968;
RA Zhou B., Berger E.G., Hennes T.;
RA "Molecular cloning of a human UDP-galactose:GlcNAc beta1,3galNAc beta1,3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme.";
RL Eur. J. Biochem. 263:571-576(1999). July 99
RP [3]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RA "Cloning and expression of two beta-1,3-galactosyltransferases: beta3gal-T5 and beta3gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RP [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-20289799; PubMed-10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann M., Rosenthal A., Kadoh T., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Saeki T., Nagamine K., Mitsuyma S., Antonarakis S.E., Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P., Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig J., Rieselmann L., Degen E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,